

Title:

Testing a computational model of rhythm perception using polyrhythmic stimuli.

Authors:

Vassilis Angelis¹, Simon Holland¹, Paul J. Upton², and Martin Clayton³

¹ The Music Computing Lab, Centre for Research In Computing, The Open University, MK7 6AA, UK

²Department of Mathematics and Statistics, The Faculty of Mathematics, Computing and Technology, The Open University, MK7 6AA UK

² Music Department, Durham University, UK

Correspondence:

Vassilis Angelis, Computing Dept, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK. E-mail: v.angelis@open.ac.uk Tel: +44 (0) 1908 33258

Title:

Testing a computational model of rhythm perception using polyrhythmic stimuli

Abstract:

Neural resonance theory (Large & Kolen, 1994; Large, 2008) suggests that the perception of rhythm arises as a result of auditory neural populations responding to the structure of the incoming auditory stimulus. Here, we examine the extent to which the responses of a computational model of neural resonance relate to the range of tapping behaviours associated with human polyrhythm perception. The principal findings of the tests suggest that: (a) the model is able to mirror all the different modes of human tapping behaviour, for reasonably justified settings and (b) the non-linear resonance feature of the model has clear advantages over linear oscillator models in addressing human tapping behaviours related to polyrhythm perception.

1. Introduction

In metrical music, the sensation of beat and meter may be viewed as a perceptual mechanism, which allows individuals to tacitly predict when in the future acoustical events are likely to occur based on an analysis of the recent musical events. One way to understand such perceptual mechanism is to approach it from a biological point of view. In this study we take the position that the biological bases of rhythm perception relate to the dynamic activity of neural populations that occurs in association with the presence of an external rhythmical stimulus. In particular, we examine a certain theory that has been suggested to account for the internal biological mechanisms of rhythm perception, namely the theory of neural resonance (Large & Kolen, 1994; Large, 2008).

This theory suggests that when a number of populations of neurons are exposed to a certain rhythmical stimulus, the firing patterns, which are driven by the structure of that stimulus, give rise to the perception of rhythm. For example, the perception of beat emerges as a result of a population of neurons firing in synchrony with the implied beat of a song. The theory of neural resonance has in turn led to the creation of a computational model (Large *et al*, 2010), which simulates the dynamic patterns of such neural activity. This computational model is a model of neural oscillation and following its mathematical derivation we refer to it throughout the paper as *the canonical model*. Briefly, a canonical model represents a dynamic system near an equilibrium state in a relatively simple form (or canonical representation) that facilitates the analysis of the dynamic system. With neural oscillation we refer to a population of both excitatory and inhibitory neurons and the properties of their interaction. The canonical model comprises a bank of tuned oscillators (i.e. each oscillator has its own natural frequency); where each oscillator accounts for modelling

a different population of neurons (i.e. neural oscillator). In the presence of a rhythmical stimulus the oscillators that best reflect the structure of that stimulus exhibit peak amplitude responses. The patterns of the amplitude responses exhibit characteristics that can reflect the multiple metric levels perceived in a rhythmical stimulus, as well as qualitative aspects of beat and meter perception.

Figure 1 illustrates an example of a dynamic pattern of the canonical model associated with the presence of a simple rhythmical stimulus, in this case a series of metronomic clicks played at 60 bpm for 50 seconds (the period of stimulation). The dynamic pattern evolves over a period of time and the diagram of *frequency response* corresponds to the average activity of each individual oscillator (blue dot) over the second half of the stimulation period (last 25 seconds). Averaging by focusing on the later parts of a stimulation period ensures that appropriate time has been given to all oscillators to respond and also, that measurements on the amplitude responses are made only after the oscillators have reached their steady states. We can see that the canonical model exhibits peak responses, which are related to the fundamental frequency of the given stimulus in three different ways: harmonically, sub-harmonically, and in a more complex fashion. More specifically, the way we calculate the frequency of the stimulus is by noting that $\text{bpm} / 60 = \text{Hz}$. For example, a series of clicks played at 60 bpm imply 1 Hz frequency, i.e., $60/60 = 1$. Consequently the fundamental frequency of the stimulus in Figure 1 is 1 Hz. From the perspective of human rhythm perception, the highest peak in Figure 1 corresponds to humans perceiving quarter notes in case of $\text{♩}=60$. Similarly, the peak of the 2 Hz oscillator corresponds to humans perceiving/ playing eighth notes, the peak of the 1.5 Hz oscillator corresponds to humans perceiving/ playing triplet crochets, the peak of the

0.5 Hz oscillator corresponds to humans perceiving/ playing half notes, and so on. Therefore, the canonical model exhibits behaviour related to the metric structure of the rhythmic stimulus, and by recalling the neural resonance theory we can say that such responses resemble the human perception of the metric levels related to a rhythmic stimulus.

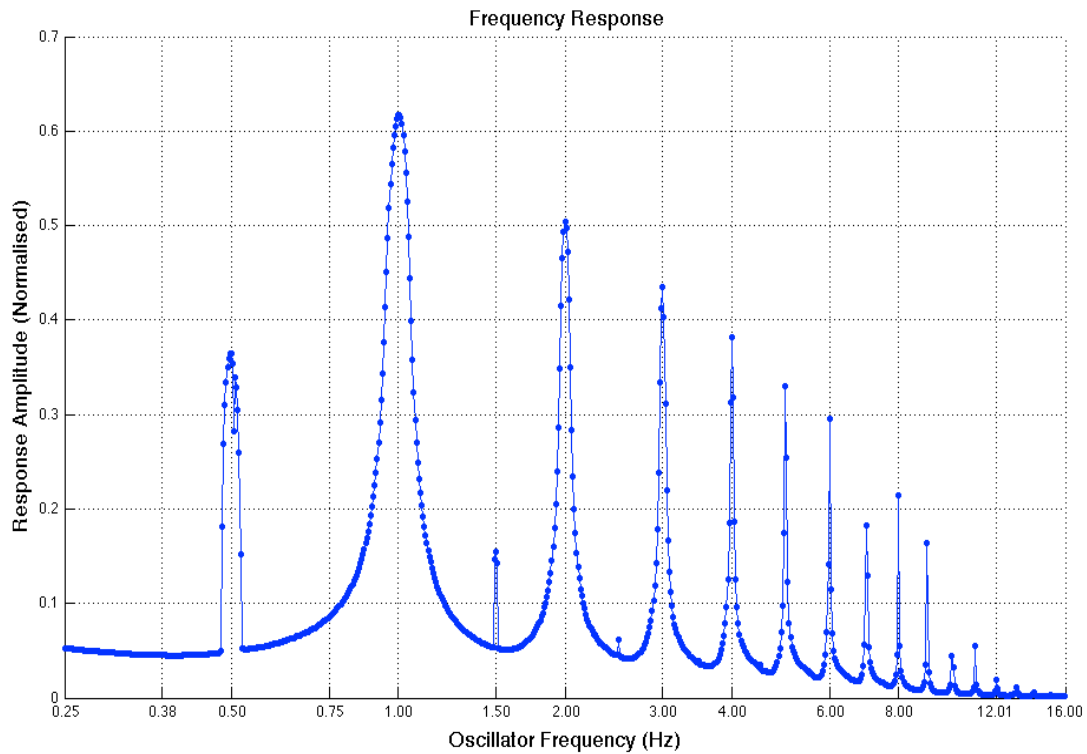


Figure 1: Average frequency response of the canonical model after being stimulated by a series of clicks played at 60 bpm. The frequency of clicks is 1 Hz, i.e. one click per second. The highest peak corresponds to the 1Hz oscillator. Peaks related to oscillators with harmonically, subharmonically and more complexly related frequencies to the 1Hz oscillator can be observed. Some of these responses can be said to resemble the human perception of different metric levels associated with a given stimulus e.g. quarter notes (1 Hz), half-notes (0.5 Hz), 16th notes (4 Hz), quarter triplets (1.5 Hz).

The number of activated oscillators and their corresponding amplitude responses comprise the two main aspects of the canonical model's behaviour as a result of being stimulated by some rhythmical stimulus. Henceforth, for purposes of concise and clear presentation, the presentation of such data (Figure 1) is described as the *frequency response* of the model in the presence of some rhythmical stimulus.

The use of polyrhythms to study aspects of human rhythm perception has been suggested (Handel, 1984) as a good experimental paradigm in balancing the complexity of the emergence of human rhythm perception encountered in musical pieces with experimental control. For example, the fact that polyrhythms have at least two conflicting pulse trains occurring simultaneously in time can be used to account for a simplified representation of the interaction of temporal, melodic, harmonic and other factors that provide periodicity information to the listener (Jones' *Joint Accent Structure Hypothesis*). Consequently, testing the canonical model using polyrhythms as stimuli provides a methodological approach for further empirical scrutiny of the theory of neural resonance and the associated canonical model.

In this paper we examine how well the canonical model can account for polyrhythm perception based on tapping behaviours reported in existing empirical studies (Handel & Oshinsky, 1981). The observed behaviours occur mainly, but not exclusively, in the form of periodic tapping along with the given polyrhythm. In other words, existing empirical experiments with humans and polyrhythms provide detailed descriptions of human behaviours associated with polyrhythm perception, therefore our main intention here is to apply to the canonical model the same polyrhythmic stimuli used with humans and investigate the extent to which the canonical model's behaviour matches aspects of human behaviour associated with polyrhythm perception. The following section describes explicitly the behavioural aspects of polyrhythm

perception, which will then provide the basis for examining the extent to which the canonical model matches human behaviour.

2. Behavioural characteristics of polyrhythm perception

Polyrhythmic stimuli in their simplest form of complexity consist of at least two different rhythms (pulse trains), which are combined to form one rhythmical structure (polyrhythmic pattern). Furthermore, “each pulse train is isochronous and unchanging, and there is a common point at which the elements of each pulse coincide” (Handel & Oshinsky, 1981). Each pulse train has a different tempo from the other, which means that the inter-onset interval (IOI) between successive events of the first pulse train is different from the IOI of the second pulse train. For example, we can think of a 3-pulse train that divides the duration of the polyrhythmic pattern in three equal IOIs, while a 2-pulse train will divide the duration of the composite pattern into two equal IOIs.

Furthermore, we should keep in mind that the polyrhythmic pattern is repeatedly presented to humans. Trivially (although vitally for clarity of argument) the repetition frequency of the stimulus (i.e. how fast the polyrhythmic pattern is repeated per unit time) can be expressed in two different ways. The first one is in terms of duration in seconds, which is what Handel and Oshinsky used. The second one is in terms of frequency, which is calculated according to the number of times the polyrhythmic pattern is repeated per second. For example, if the polyrhythm repeats itself once per second then its period expressed in frequency terms will be 1Hz. This second convention is useful when it comes to encode the different repetition rates (used by Handel & Oshinsky) in a form that is recognisable by the canonical model (see Table 1).

2.1 Range of tapping behaviours

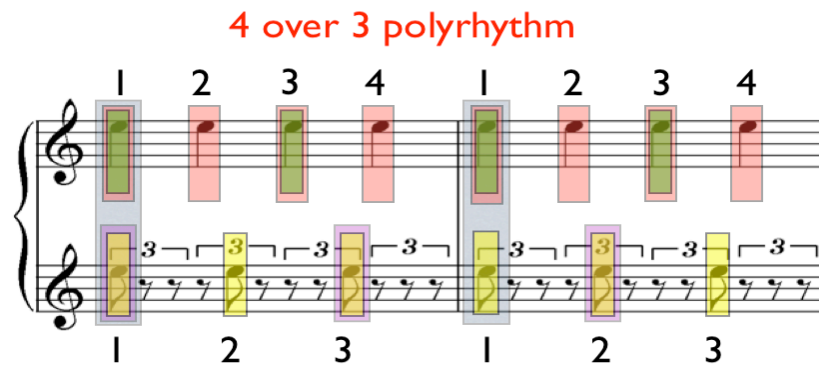
The way humans perceptually organise a polyrhythm is reflected by the way they tap along with a given polyrhythm and the relative preferences of such tapping behaviours. Considering a polyrhythm of two rhythms m and n , the most frequently occurring ways of perceptually organizing and therefore tapping along with a given polyrhythm are (Pressing *et al*, 1996):

- (a) Tapping along with the composite pattern of m and n rhythms.
- (b) Tapping along with the m rhythm.
- (c) Tapping along with the n rhythm.

Additionally, some less frequent modes of tapping along with a polyrhythm include (Handel & Oshinsky, 1981):

- (d) Tapping along with the points of co-occurrence of the two rhythms (i.e. once per pattern repetition or unit meter),
- (e) Tapping along with every second element of one of the rhythms, e.g. most commonly tapping every other element of a 4-pulse train and less commonly every other element of a 3-pulse train.

Figure 2 below illustrates cases b to e mentioned above for a 4:3 polyrhythmic stimulus. The case of a 4:3 polyrhythm is a good representative of tapping behaviours for a range of two pulse-train polyrhythms (e.g. 3:2, 5:2, 3:5, 4:5) so it is used here as a workhorse example in comparing human tapping behaviours with the frequency response of the canonical model.



Range of human tapping behaviour to be explained

- Tapping along with the 4-pulse rhythm
- Tapping along with the 3-pulse rhythm
- Tapping along with the coincidence of the two rhythms
- Tapping along with every other element of the 4-pulse rhythm
- Tapping along with every other element of the 3-pulse rhythm

Figure 2: Range of periodic tapping behaviours along with a 4:3 polyrhythm

In general, according to Handel & Oshinsky (1981), 80% of the behaviourally observed responses correspond to tapping along with the elements of one of the pulse trains (red and yellow bands). The second major class (12% of the observed behaviours) of tapping along with a polyrhythmic stimulus in a periodic way, is to tap in synchrony with the co-occurrence of the two pulses, or once per unit meter (grey band). The third class (6%) concerns humans tapping periodically along with every other element of a 4-pulse train (most common), and every other element of a 3-pulse train (least common) (green and purple bands). We should note that cases (d) and (e) of the list imply a submultiple relation to the two main pulse trains, which means that humans are able to tap in a periodic way by concentrating onto some events while ignoring others.

One useful step to facilitate the comparison between the tapping behaviours listed above and the frequency response of the canonical model is to express the former in frequency terms. Remember that the way the canonical model responds in the

presence of a rhythmical stimulus is to resonate at frequencies related to the stimulus (Figure 1), therefore by expressing the tapping behaviours in frequency terms we can directly compare them with the frequency response of the model. Table 1 below expresses the tapping behaviours listed above on the basis of a 4:3 polyrhythmic stimulus that repeats once per second, i.e., the repetition frequency of the pattern is 1 Hz. Note that case (a) of the list is ignored, as this tapping behaviour is not isochronous.

Tapping along with the 4-pulse train.	4Hz
Tapping along with the 3-pulse train.	3Hz
Tapping along with the co-occurrence of the two pulses.	1Hz
Tapping along with every second element of the 4-pulse train.	2Hz
Tapping along with every second element of the 3-pulse train.	1.5Hz

Table 1: Tapping behaviours along with a 4:3 polyrhythmic stimulus expressed in terms of frequencies for a 1Hz frequency repetition of the polyrhythmic pattern.

Before we move onto the next section we consider the response of a series of linear oscillators in the presence of the above 4:3 polyrhythmic pattern that repeats once per second as a reference comparison with the above human tapping behaviours. Figure 3 below illustrates the frequency response of a bank of linear oscillators in the presence of the 4:3 polyrhythmic stimulus. In comparison with human tapping behavior, it is worth noting that there is no response at all corresponding to cases (d) and (e) of the human tapping behavior.

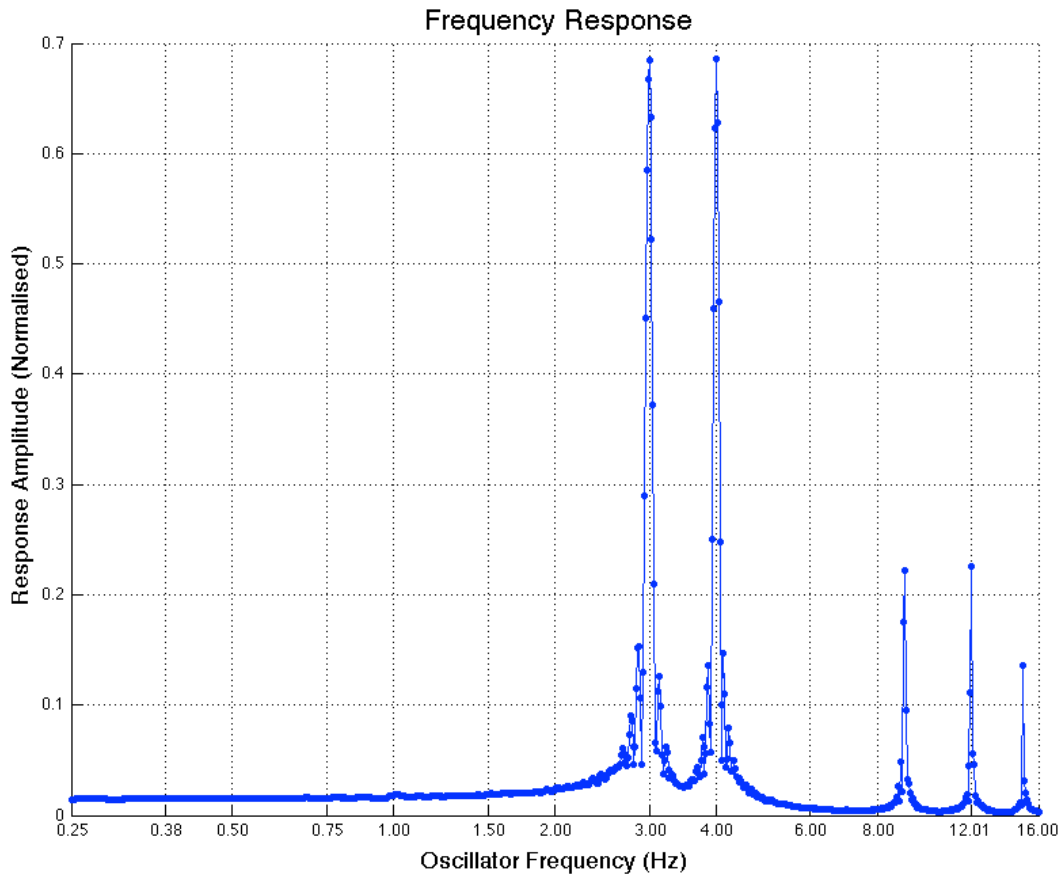


Figure 3: Response of a series of linear oscillators in the presence of a 4:3 polyrhythmic stimulus repeating once per second. The polyrhythm consists of two square waves 4Hz and 3Hz.

3. Method

In this section we provide background information about the conducted tests, and we explain the extent to which certain settings of the canonical model reasonably correlate to aspects of human auditory physiology. However, this paper does not include an analysis related to the parameters of the actual state variable of the oscillator. Such analysis is under preparation as part of the uncertainty analysis of the hereby-presented results using parameter sensitivity analysis and comparison with other computational models. The canonical model is a bank of oscillators tuned to different frequencies, which are arranged from low to high frequency. The oscillators have been designed to be non-linear, based on evidence that suggests that the auditory

nervous system is highly nonlinear, and that nonlinear transformations of auditory stimuli have important functional consequences (Large *et al*, 2010). With the aforesaid assertion in mind we expect that the non-linear transformations of the polyrhythmic stimulus produced by the model would resemble behavioural aspects of polyrhythm perception. In such a case, evidence that this was indeed the case would support the thesis that polyrhythm perception and its associated behaviours are indeed related to non-linear transformations of the polyrhythmic stimulus in the auditory nervous system. Here we consider the following components related to setting up the canonical model for the purposes of conducting the tests:

- a) Input encoding
- b) Frequency range of the bank of oscillators
- c) Number of oscillators
- d) Duration of stimulation
- e) Connectivity of oscillators and number of networks

3.1 Input encoding

The essential idea for encoding and presenting a rhythmical stimulus to the canonical model in our evaluation is to recreate a stimulus similar to the one presented to humans in the Handel & Oshinsky experiment. When presented to humans a simple polyrhythmic stimulus consists of two individual pulse trains, each of which is delivered through a different loudspeaker placed in front of the subjects. The physical characteristics of the stimulus, such as the amplitude of each pulse train, the pitch of each pulse train and the presentation tempo, are experimental variables. For example, in experiments where the focus is on examining the influence of the tempo in

polyrhythm perception, both the amplitude and the pitch of each pulse train are kept the same.

In our evaluation, the polyrhythmic stimulus we used to stimulate the canonical model is intended to mirror that which Handel and Oshinsky (1981) used in their experiments. In this way, we can use as a reference empirical data describing the ways humans tap along with a given polyrhythmic stimulus (Section 2), and we can make comparisons with the way the canonical model responds in the presence of the same stimulus (Section 4).

In order to test the canonical model we are faced with a technical choice of encoding a polyrhythmic stimulus using any of three ways: functions, audio files sampling the stimulus and MIDI files representing times of events. In this series of tests we have created audio files to represent the polyrhythmic stimulus. Additional tests using both of the alternative methods mentioned above have been undertaken with no major effect in altering the results presented here. The polyrhythmic stimulus was created in Matlab as a combination of two square-waves. Section 3.3 provides more details on the exact pairs of frequencies we used to compose the 4:3 polyrhythmic stimulus for a number of different repetition rates similar to those used by Handel & Oshinsky's empirical study (Table 2).

Ideally, we would like to encode the stimulus to exactly match the stimulus used in human experiments. This would mean that we would have to consider the physical characteristics of the stimulus presented in humans, such as its amplitude, its pitch, its ADSR envelope and its tempo. When using a square-wave function to encode the stimulus there is a limitation in incorporating all the above characteristics, because the square-wave function resembles pulses rather than tones, and as a result some

characteristics are compromised. However, as long as the current version of the canonical model considers only the temporal characteristics of the stimulus, the aforementioned limitation does not create any obstacles.

3.2 Frequency range of the bank of oscillators

In order to define the range of natural frequencies of the bank of oscillators we have taken into consideration two things. The first one is the range within which humans are able to perceive beat and meter and its relation to the observed tapping frequencies in polyrhythm perception, and the second one is to make sure that there is an individual oscillator with natural frequency in the bank that matches any potential tapping frequency observed in experimental data. More specifically, the first point will help us to define the limits of the range, while the second one will guide us towards the granularity of the range. Finally, we define the former parameter by reviewing the literature (this section) and the latter, by considering qualitative characteristics of the nature of the oscillators to address the degree of granularity needed (Section 3.3).

The perception of beat and meter implies that events can be perceived as being both discrete and periodic at the same time. The lower inter-onset interval (IOI) for perceiving two events as being distinct is 12.5 msec (Snyder & Large, 2005, p.125). However, humans can start perceiving (or tracking) that a series of events is indeed regularly periodic, only if the IOI between successive events is at least 100 msec. According to London (2004), the 100 msec interval refers to the shortest interval we can hear or perform as an element of a rhythmic figure. This limit is well documented in the literature as the lower limit for perceiving the potential periodicity of successive events. (London, 2004; Snyder & Large, 2005; Pressing & Jolley Rogers, 1997; Repp,

2005). In the Handel & Oshinsky empirical observations the lowest limit observed is tapping along with a 3-pulse train played at 600 msec, which implies tapping once every 200msec.

Regarding the upper limit for perceiving periodic events there are several suggestions about how long it can be, and as Repp says, it is “a less sharply defined limit” (2005). For example, London suggests that the longest interval (upper limit) we can “hear or perform as an element of rhythmic figure is around 5 to 6 secs, a limit set by our capacities to hierarchically integrate successive events into a stable pattern” (2004, p.27). Repp quotes Fraisse’s upper limit, which is said to be 1800 msec (2005), and in a more recent study (2008) he points out to the difficulty in providing evidence towards a sharply defined upper limit of sensorimotor synchronisation tasks (SMS). In this study we are mainly interested in capturing tapping frequencies that are observed in Handel and Oshinsky’s study and therefore we consider the upper limit with regard to these observations. In particular, we considered tapping behaviours when the experimental variable was the tempo of the polyrhythm exclusively. In that case the upper limit is tapping along with the unit pattern of the polyrhythm for a repetition rate of 1 second, therefore 1000 msec.

We can define the natural frequencies of the oscillators of the canonical model in a way that is arbitrary, but nonetheless informed by the following considerations:

- the limits observed in SMS studies (as noted above),
- the tapping frequencies observed in Handel & Oshinsky’s study,
- potential tapping frequencies related to plausibly perceivable repetitive patterns in a polyrhythmic stimulus.

Just as in the earlier discussion of the range of tapping behaviours related to polyrhythms, for convenience it helps to express the SMS range in terms of

frequencies by making use of the $F=1/T$ formula, where F is the frequency and T is the “tapping” period in seconds. For example, tapping once every 0.1 sec (100 msec) implies a tapping frequency of 10 Hz, while tapping once every 2 seconds (2000 msec) implies a frequency of 0.5 Hz. Table 2 below illustrates potential tapping frequencies associated with the empirical study of Handel and Oshinsky. The repetition rate of the polyrhythmic pattern is given in terms of seconds in the first column, and it has been translated in frequency terms in the second column (rounded to 2 d.p where necessary), again by using the $F=1/T$ formula. Similarly, based on the unrounded frequency of the pattern’s frequency we calculated the frequencies for the rest of the columns and rounded to 2 d.p where necessary. For example, when the repetition rate is 2.4 secs, the frequency of the pattern is $1/2.4 = 0.4166666667$ Hz, which is rounded up to 0.42 Hz. To calculate the fundamental of the 4-pulse train, we multiplied the unrounded pattern’s frequency (0.4166666667 Hz) by 4, which equals to 1.666666667 Hz, and we rounded it up to 1.67 Hz.

Repetition Rate in sec	Pattern’s Frequency	Fundamental of 4-pulse train	1st sub-harmonic of 4-pulse	Fundamental of 3-pulse train	1st sub-harmonic of 3-pulse
2.4	0.42Hz	1.67Hz	0.83Hz	1.25Hz	0.63Hz
2.0	0.50Hz	2.00Hz	1.00Hz	1.50Hz	0.75Hz
1.8	0.56Hz	2.22Hz	1.11Hz	1.67Hz	0.83Hz
1.6	0.63Hz	2.50Hz	1.25Hz	1.88Hz	0.94Hz
1.4	0.71Hz	2.86Hz	1.43Hz	2.14Hz	1.07Hz
1.2	0.83Hz	3.33Hz	1.67Hz	2.50Hz	1.25Hz
1.0	1.00Hz	4.00Hz	2.00Hz	3.00Hz	1.50Hz
0.8	1.25Hz	5.00Hz	2.50Hz	3.75Hz	1.88Hz
0.6	1.67Hz	6.67Hz	3.33Hz	5.00Hz	2.50Hz

0.4	2.50Hz	10.00Hz	5.00Hz	7.50Hz	3.75Hz
-----	--------	---------	--------	--------	--------

Table 2: Potential tapping behaviours along with a 4:3 polyrhythm expressed in frequencies for a series of different repetition rates.

Once we decide on the frequency range we want to incorporate into the canonical model based on the above discussion (e.g. 0.42 Hz to 10Hz), the means of executing this is directed by the way in which the software that implements the canonical model (Large et al, 2010) is currently designed. One way of doing this is by defining a certain central frequency and then deciding how many octaves need to be added either side of that central frequency, together with how many oscillators are required within each octave (Section 3.3). So for example if we want a frequency range that covers a tapping frequency range (0.5Hz to 10Hz) we can set the central frequency of the bank at 2Hz and then we could add three octaves on either side, i.e. six octaves in total. In this case, the frequency range will be 0.25 to 16Hz.

3.3 Number of oscillators

As we stated in Section 3.2, in order to define the granularity of the frequency range we have to make sure that for each empirically observed or potential tapping frequency there is an oscillator with natural frequency quite close to it to avoid bias in the amplitude responses. The nonlinear oscillators in the model can have extremely high frequency resolution, and this resolution depends on the amplitude of the stimulus. Therefore, we have to make sure that the oscillators are packed tightly enough in frequency space in order to respond to all frequencies. A large number of oscillators per octave such as 128 provides a sufficient level of granularity as needed.

3.4 Duration of stimulation

The duration of stimulation of the model should be long enough in order for the oscillators to reach a steady state. Steady states can be observed by obtaining spectrograms covering the stimulation period and spotting a point in time at which the systems appear stable. Figure 4 below illustrates a stimulation period of 6 seconds using a 4:3 polyrhythm which is repeated once every second. The y-axis represents the frequency range of the oscillators (0.25Hz-16Hz, see section 3.2). The redness indicates energy levels and in effect shows which oscillators are mostly activated in the presence of this particular polyrhythmic stimulus. For example the oscillators with 3Hz and 4Hz natural frequencies concentrate the largest amount of energy, followed by oscillators, which are harmonically and subharmonically related to the aforementioned frequencies. The x-axis represents time, which manifests the overall dynamics of the system over the time of stimulation. This graph allows us to illustrate the importance of allowing sufficient time for the system to reach a steady state and also the importance of averaging the activity only after a steady state has been reached. For example if we had only considered a 2-second stimulation period and averaged over 100% of that period we would have observed dynamics over the transient phase of the system. In that case important information about the sub-harmonic responses of the system with regard to the fundamental frequencies of the stimulus may have been lost, which is a crucial piece of information for the arguments we are trying to make.

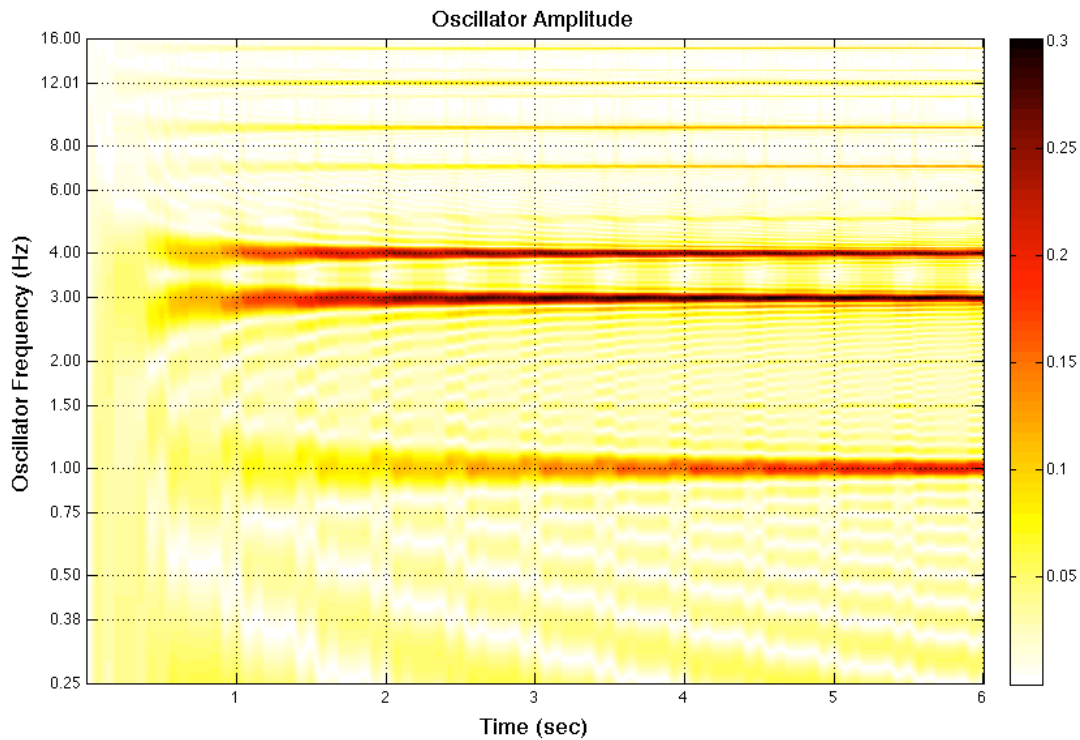


Figure 4: Oscillators' amplitude response to a 4:3 polyrhythmic stimulus. The polyrhythmic pattern repeats once every second over a period of 6 seconds.

In the Handel & Oshinsky study the polyrhythmic stimuli were initially presented for 15 seconds, followed by a 5 seconds silent pause after which the subjects were asked to start tapping. However, the exact time it takes before subjects start tapping is not documented in that paper. In another study, the time to start tapping along with a given rhythmical stimulus was reported (Snyder & Krumhansl, 2001). In particular, they report that four beats are needed before start tapping (BST), which they suggest might typically correspond to approximately 2400 msec. Large (2000a) suggests that the BST time corresponds to the amount of time needed to reach a stable limit cycle with regard to the dynamics. In Figure 4 we can see that the timescale for the oscillators to reach a stable limit cycle is similar to the time reported by Snyder & Krumhansl. Consideration of Figure 4 suggests that the time before start tapping (BST) would vary for different tapping frequencies along with the polyrhythmic

stimulus. This opens up an opportunity for obtaining new predictions about human behaviour from the model and thus testing it further.

3.5 Oscillators' connectivity and number of networks

In principle, we can have the oscillators of the bank interconnected to each other in order to form a network of oscillators, which is believed to be a more accurate representation of the underlying connectedness of neural populations in the auditory nervous system. However, in this particular series of tests, partly for reasons of simplicity, we examine the individual responses of each of the oscillators to the external polyrhythmic stimulus, i.e. there is no internal coupling among the oscillators. This chosen arrangement “focuses the response of the networks to the external input” (Large, 2010, p.4). As a further choice in setting up the model, we could choose to have more than one bank of oscillators. One argument in favour of assuming more than one bank of interconnected oscillators (network of oscillators) is related to the fact that processing of any auditory stimulus takes place in more than one area in the brain, thus utilising more than one network would be more plausible based on the above physiological basis. As Velasco suggests (personal communication), those two networks might be small patches of tissue in the same local brain area, or they could be two areas that are far away from each other, for instance, primary auditory cortex and supplementary motor area (SMA).

Taking into account the above considerations, in the conducted tests we have nevertheless employed a single network of non-interconnected oscillators, which is the simplest option for starting to test the canonical model. In effect, the simplified version of the model puts the focus on its non-linear resonance feature. Further

investigations involving both interconnections and more than one network are planned as part of our future work.

4. Results

Recall that the 4:3 polyrhythm we have chosen with which to stimulate the canonical model can be viewed as a good focus for comparing human and model behaviour since it elicits a sufficient range of types of human tapping behaviours observed in a series of simple polyrhythms like 3:2, 2:5, 3:5, 4:5. The polyrhythmic stimulus, encoded along the lines described in Section 3.1, was used to stimulate the canonical model for a sufficient period (i.e. providing time for the system to reach steady state) at ten different repetition rates. These rates were exactly the same as the ones used by Handel and Oshinsky's experiment, with some approximations as discussed (Table 2).

The conclusions we draw in the discussion section are based on the analysis of the results for all ten repetition-rates of the polyrhythmic stimulus. Briefly, the two additional figures (a slower and a faster rate compared to the 1Hz) presented in the Appendix A share the same basic shape apart from being shifted across the frequency axis. Therefore, the *frequency response* for just one of these rates is sufficient to give the context needed to understand the results. Additionally, results from different type of polyrhythms such as those mentioned above e.g. 3:2 and 3:5 are included in the Appendix B.

Figure 5 below shows the averaged amplitude responses of the canonical model in the presence of a 4:3 polyrhythmic stimulus for the last 20% of the stimulation period. The spectrum analysis of the system's response indicated that over the last 20% of the stimulation period the system has reached its steady state, therefore we averaged over that period. For facilitating illustration we have chosen the polyrhythmic pattern that

repeats once every sec. Each amplitude peak in the *frequency response* graph below results from the activation of a series of oscillators. However, there is one particular oscillator that exhibits the peak amplitude and we refer to it as the *main* oscillator. Additionally, in Figure 5 some amplitude peaks have been labelled using a green colour in order to indicate that they correspond to human tapping behaviours. More specifically, the main oscillator of the green-labelled peaks has a natural frequency that reflects a human tapping frequency. We will analyse these results in the discussion section in detail. Below we list the principal peaks in these figures that can be correlated with human tapping behaviour.

- 1) A peak with a main oscillator of 3 Hz natural frequency.
- 2) A peak with a main oscillator of 4 Hz natural frequency.
- 3) A peak with a main oscillator of 1 Hz natural frequency.
- 4) A peak with a main oscillator of 2 Hz natural frequency.
- 5) A peak with a main oscillator of 1.5 Hz natural frequency.

Additionally, significant peaks corresponding to oscillators with natural frequencies harmonically related to the frequencies of the main oscillators listed above (e.g. 3rd harmonic of the 3Hz and 4Hz oscillator, 5th and 7th harmonic of the 1Hz oscillator) were also observed. However, these peaks do not appear to directly correspond to any human behaviour observed in the Handel & Oshinsky study (red colour-code). Nonetheless, some of these harmonic responses, such as 5Hz, 6 Hz, and 7 Hz (see 3.2), correspond to attainable tapping frequencies by humans.

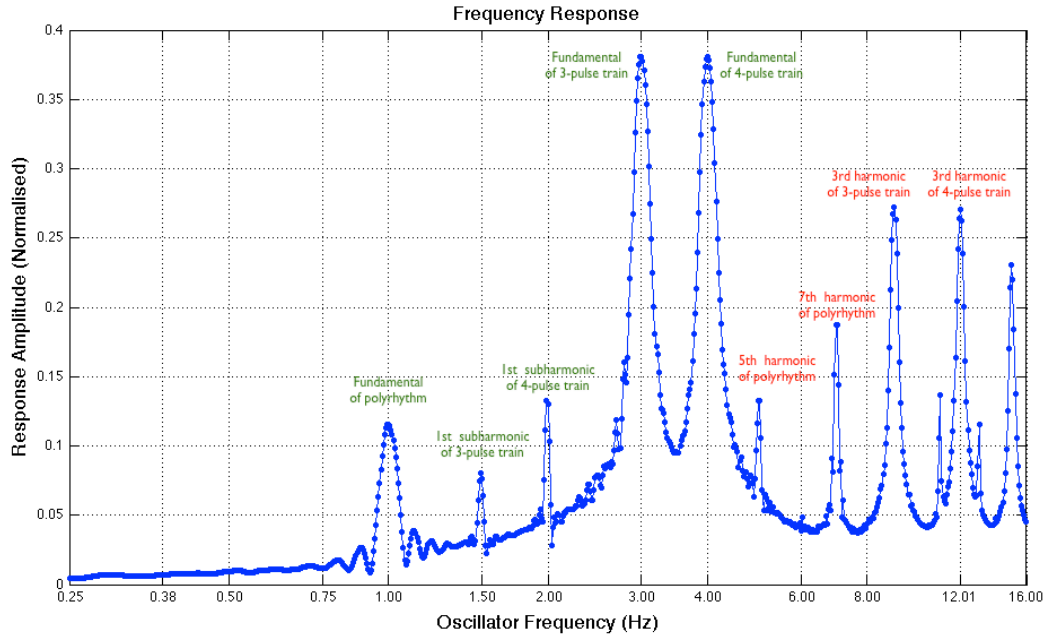


Figure 5: Frequency response in the presence of a 4:3 polyrhythm. The polyrhythmic pattern is repeated once every sec for about 80 secs. The figure illustrates the average amplitude over the last 20% of the stimulation. The *main* oscillators in the labelled green peaks represent oscillators with natural frequencies similar to the human tapping frequencies (e.g. fundamental of polyrhythm, 1st sub-harmonic of the 3-pulse train, 1st sub-harmonic of the 4-pulse train or 2nd harmonic of the polyrhythm, fundamental of the 3-pulse and 4-pulse train). The red-labelled peaks are formed by oscillators with natural frequencies that correspond to harmonics of both pulse trains (e.g. 3rd harmonic of both 4-pulse and 3-pulse train, and harmonics of the polyrhythm’s fundamental frequency), and more odd overtones.

Before we proceed with the discussion of the results we should note that tapping every other element of a 4-pulse train corresponds not only to the 1st sub-harmonic of the fundamental frequency of the 4-pulse train but also to the 2nd harmonic of the fundamental frequency of the polyrhythm. We can therefore assume that the final amplitude response is a combination of both.

5. Discussion

In this section we discuss the extent to which the peaks produced by the canonical model at various tempi account for the *variety* of human tapping frequencies. Also, we discuss different transformation methods for analysing the polyrhythmic signal in

order to point to the structural components of a stimulus and their potential effect in rhythm perception.

As we briefly mentioned in section 4, the canonical model resonates at frequencies, which can be compared with the way humans interpret polyrhythmic stimuli by tapping along with a given stimulus in a periodic way. More specifically, we have found that of the five observed modes of human periodic tapping to a 4:3 polyrhythmic stimulus (the five cases displayed in Figure 2), the canonical model predicted all five for all repetition rates.

The extent to which the variety of human behaviours corresponds with the peaks exhibited by the canonical model is summarised in Table 3. The table shows the results for 1Hz repetition frequency of the polyrhythmic stimulus.

	Human behaviours	Freq (Hz)	Peaks
1	Composite pattern $m+n$	N/A (irregular)	N/A (Irregular)
2	m rhythm	4	Fundamental of 4-pulse train
3	n Rhythm	3	Fundamental of 3-pulse train
4	Unit pattern	1	Fundamental of polyrhythm
5	Every second element of the 4 pulse train	2	2 nd harmonic of polyrhythm & 1 st subharmonic of 4-pulse train
6	Every second element of the 3 pulse train	1.5	1 st subharmonic of 3-pulse train

Table 3: Comparison of human tapping behaviours with model's frequency response to a 4:3 polyrhythmic stimulus repeating once per second.

An interesting point regarding the nature of the canonical model is the subharmonic responses (e.g. cases 4-6 in Table 3) it produces in relation to the fundamental frequencies of the two pulse trains of the polyrhythmic stimulus. For example, there is no pulse-train explicitly present in the initial stimulus of the polyrhythm with exactly the same frequency as the one implied by the canonical model's response regarding the repetition frequency of the polyrhythm. Such subharmonic responses can be attributed to the non-linear features of the model in the following sense. Figure 3 of Section 2.1 shows the response of a model consisting of series of linear oscillators to the same 4:3 polyrhythmic stimulus we used in the canonical model as well. The linear model exhibits no subharmonic responses. But such kind of linear model is essentially the canonical model with the non-linear features switched off. Thus, the subharmonic responses may be attributed to the non-linear features of the canonical model.

When comparing candidate formal models for human rhythm perception, it is useful to be aware of two different view-points on periodic phenomena, namely frequency and periodicity. Frequency involves wave phenomena (e.g. audio tones) and can be exhaustively analysed without loss of information by Fourier analysis. By contrast, periodicity requires only temporal sequences of point-like events (e.g. rhythms) and can be analysed by a variety of pattern recognition techniques. Diverse methods such as continuous wavelets (Smith & Honing, 2008), autocorellation and periodicity transforms (Sethares, 1999) may recognise different kinds of repeated patterns in time series stimuli. Despite the clear distinction between these two viewpoints, any periodic sequence or combinations of periodic sequences like the 4:3 polyrhythm, can always be represented by a waveform (e.g. an appropriate square wave) and subjected to Fourier analysis. Fourier analysis then provides the last word on what frequency

components are present ‘in the signal’. However, additional repeating patterns (e.g. subharmonics in polyrhythmic patterns) may be recognizable by different methods that may go beyond frequency components. Thus, approaches such as continuous wavelets (Smith & Honing, 2008), periodicity transforms (Sethares, 1999), and autocorrelation may be able to produce responses that relate to the subharmonic responses discussed above. However, the canonical model not only matches such human responses, but also has the advantage of close ties with neurological theory and has physiological plausibility.

As noted above, Fourier analysis exhaustively analyses, in a well-defined sense, the frequency components of a stimulus. Figure 6 is a time series representation of the 4:3 polyrhythmic stimulus, which is created as a composition of two square waves of 4 Hz and 3 Hz.

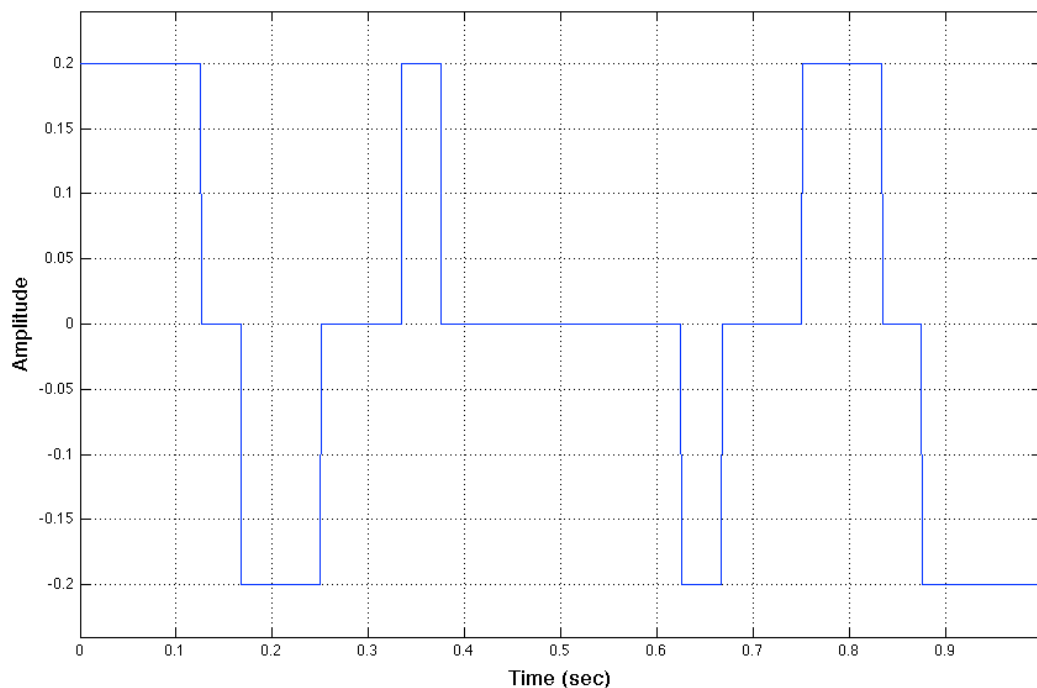


Figure 6: Time series representation of the 4:3 polyrhythmic stimulus. The stimulus is a combination of two square waves of 4 Hz and 3 Hz. The graph shows time series over one cycle of completion of the polyrhythmic pattern.

Figure 7 below illustrates the result after an FFT analysis has been applied to a 4:3 polyrhythmic stimulus with repetition frequency of 1Hz. The frequency spectrum has been limited to match the 0.25 – 16 Hz range used with the canonical model. The figure exhaustively identifies the fundamental frequency components of the stimulus, and the odd harmonics of the fundamentals (inasmuch the fundamentals are squarewaves) as expected. However, different transformations may treat the polyrhythmic signal as a time series data, and in such cases patterns can be identified that relate subharmonically to the fundamentals.

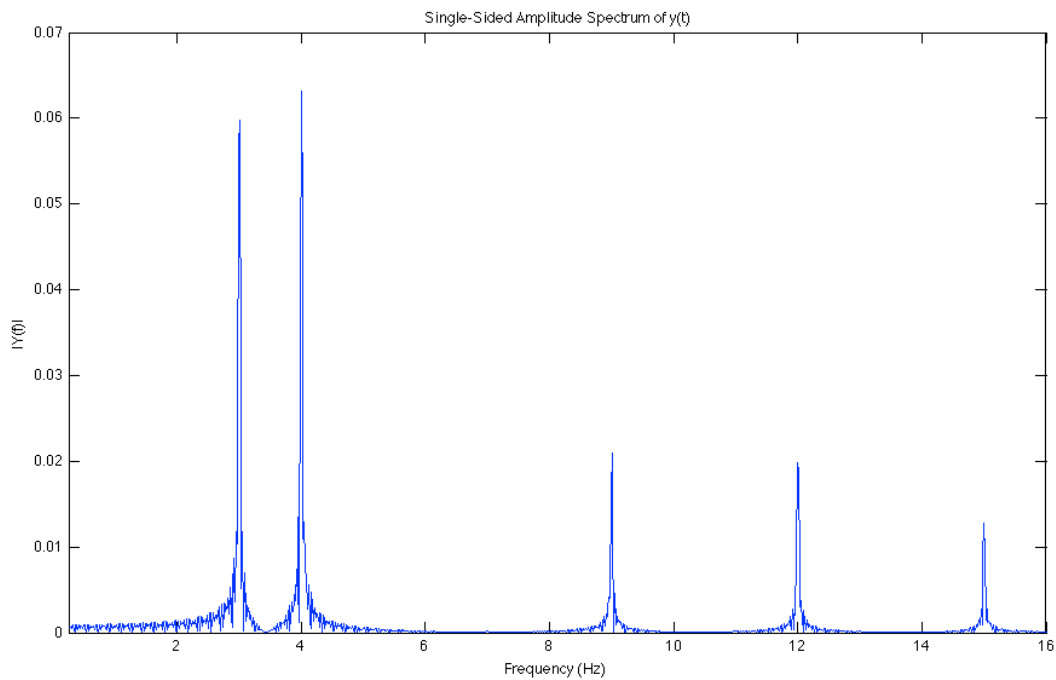
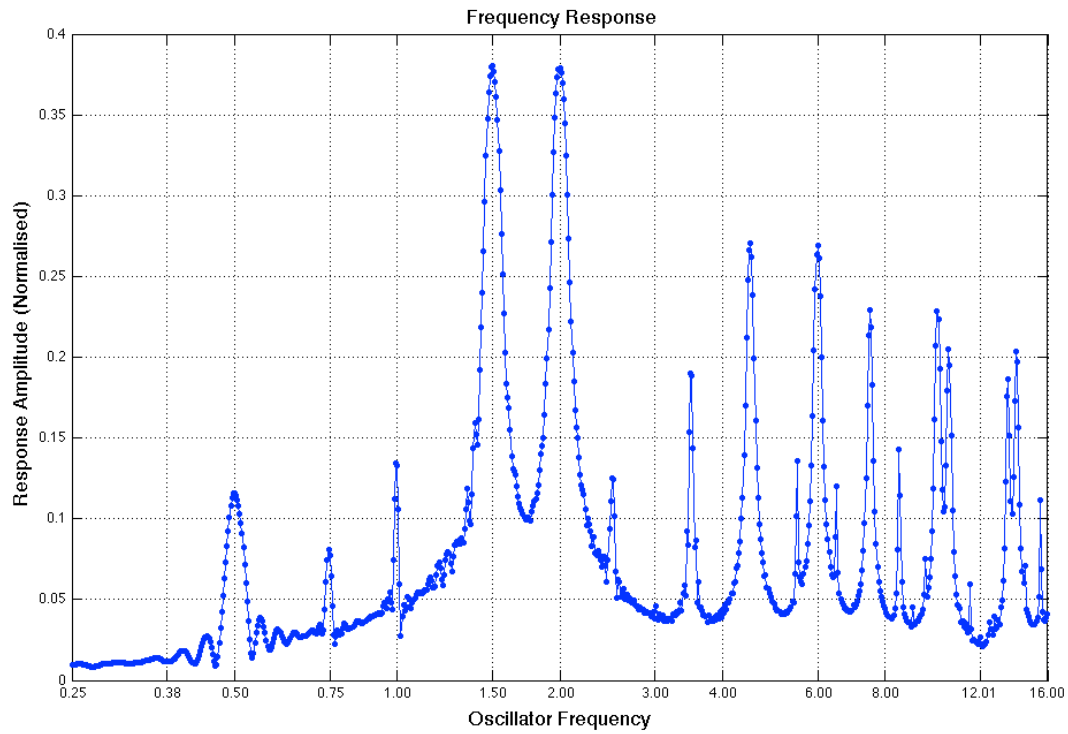


Figure 7: FFT analysis of a polyrhythmic signal comprising of two square waves, one 4Hz and one 3Hz.

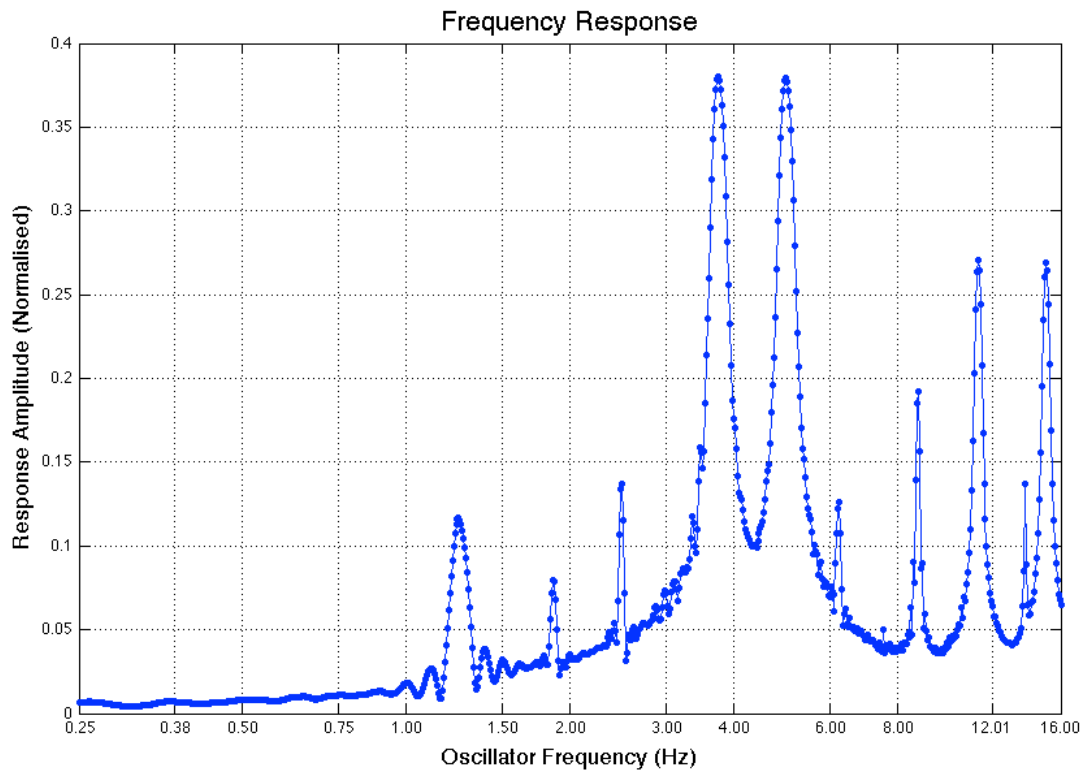
6. Conclusions

In this paper we have explored the extent to which the canonical model and its role as a model of rhythm perception could match aspects of polyrhythm perception. The canonical model is an instantiation of the neural resonance theory and its fundamental units are non-linear oscillators able to resonate in the presence of some rhythmical stimulus. The canonical model provided responses, which address the full range of tapping behaviours encountered in the particular case of a 4:3 polyrhythm. We have also illustrated the importance of the non-linear nature of the model in capturing the aforementioned tapping behaviours by comparing its responses to a linear model of a series of linear oscillators and also. Finally, if the theory of neural resonance provides a good account of the reaction of populations of neurons in the presence of some rhythmical stimuli, this paper suggests that it is the nonlinear transformations of polyrhythmic stimuli in the brain of humans, which are partially responsible for the overt tapping behaviours in polyrhythm perception.

APPENDIX A: Responses of the canonical model with regard to two additional repetition rates (2000 msec and 800 msec) of the 4:3 polyrhythmic stimulus, which are encountered in the empirical study of Handel & Oshinsky (1981).

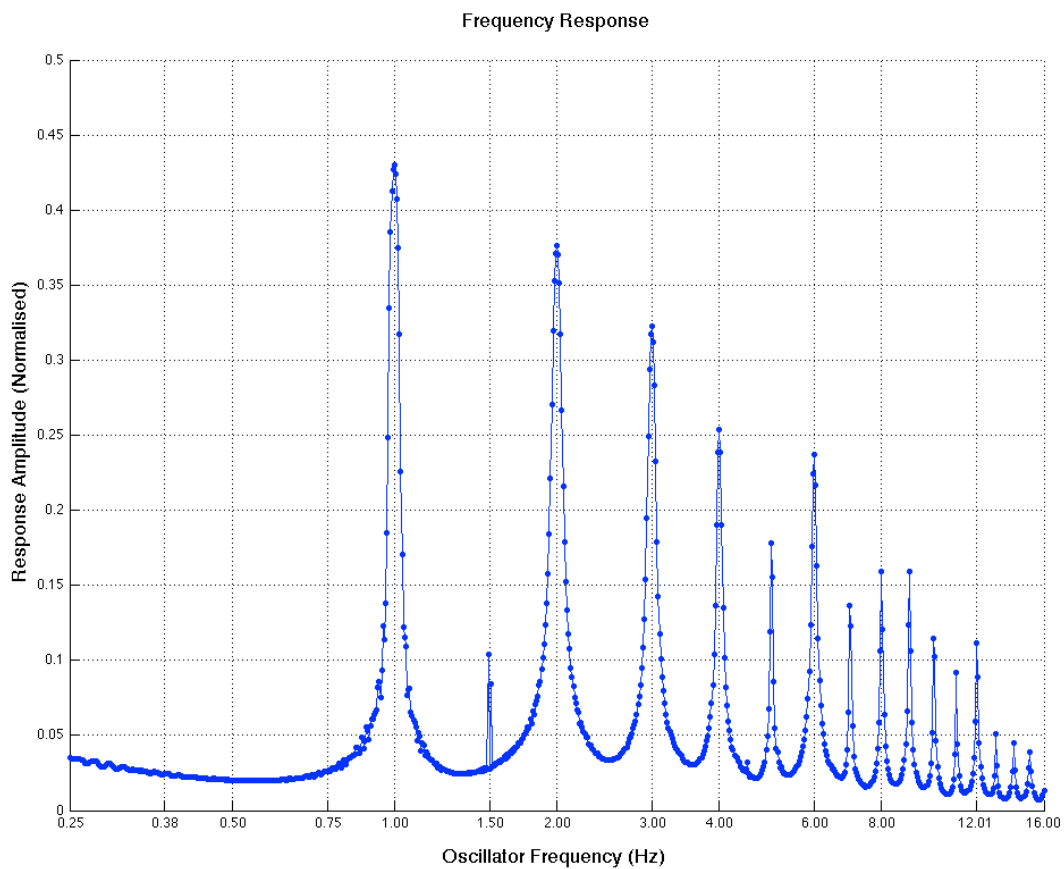


Frequency response of the canonical model in the presence of a 4:3 polyrhythm repeating once every 2000 msec.

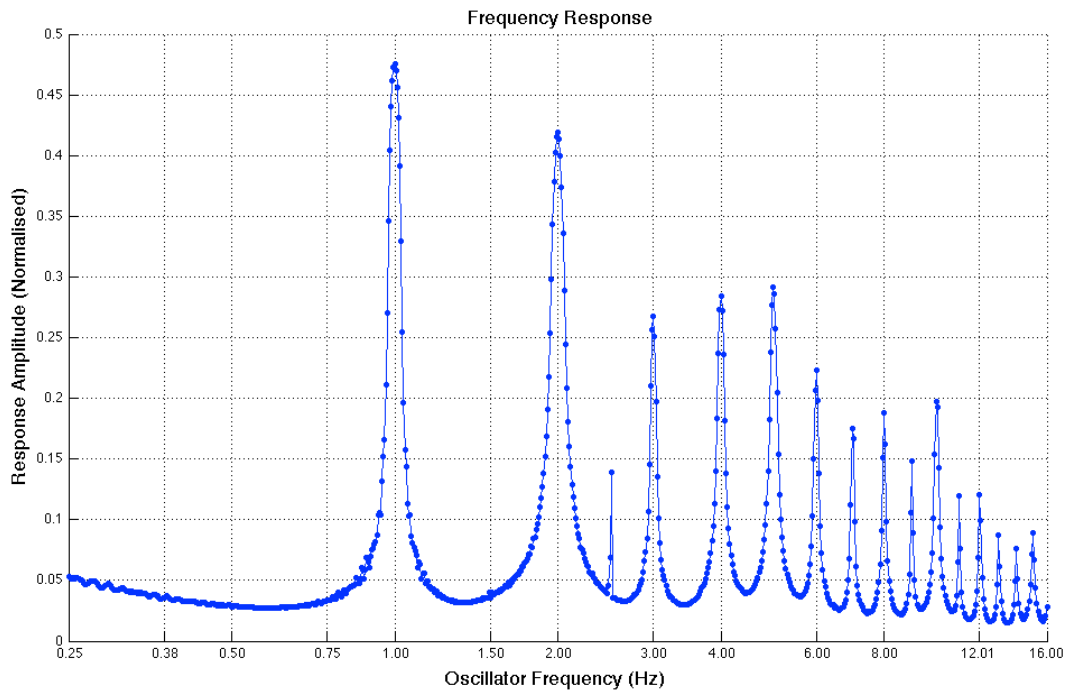


Frequency response of the canonical model in the presence of a 4:3 polyrhythm repeating once every 800 msec.

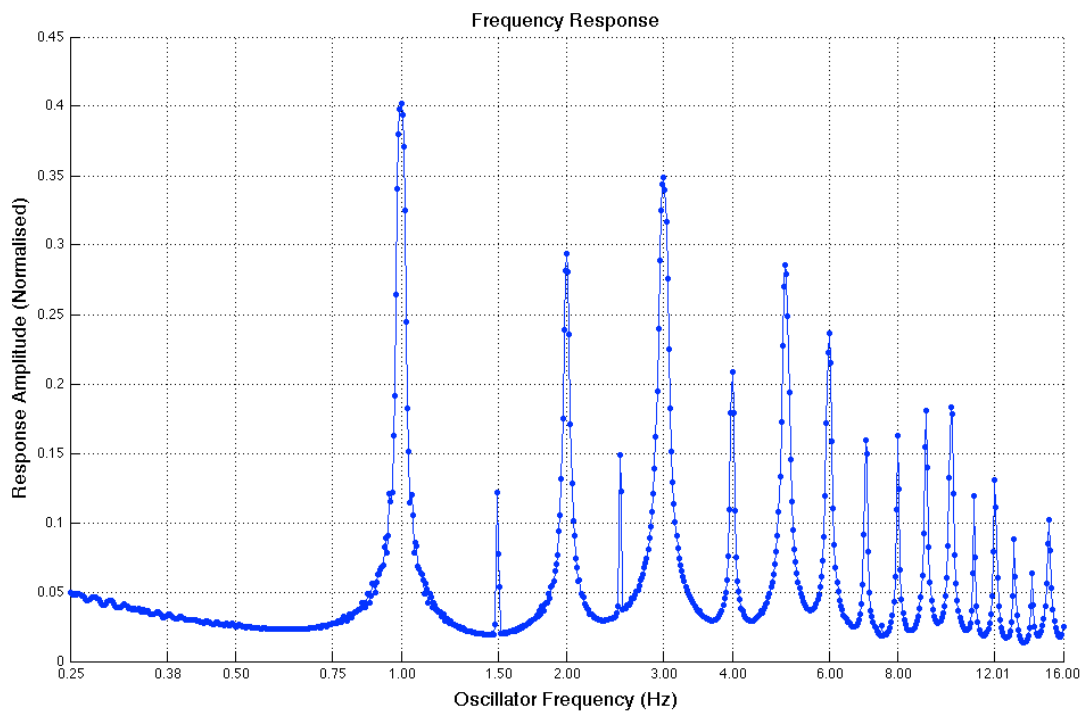
Appendix B – Frequency response of the canonical model in the presence of other types of two-pulse train polyrhythms such as 3:2, 2:5, 3:5, 4:5. Note that responses related to the first sub-harmonics of the fundamentals and the unit meter are present. For facilitating presentation the repetition rate of the pattern for all polyrhythms has been chosen to be 1 repetition per sec, i.e. 1 Hz. In that case the above notation of the polyrhythms also implies the frequency of the fundamentals (pulse trains).



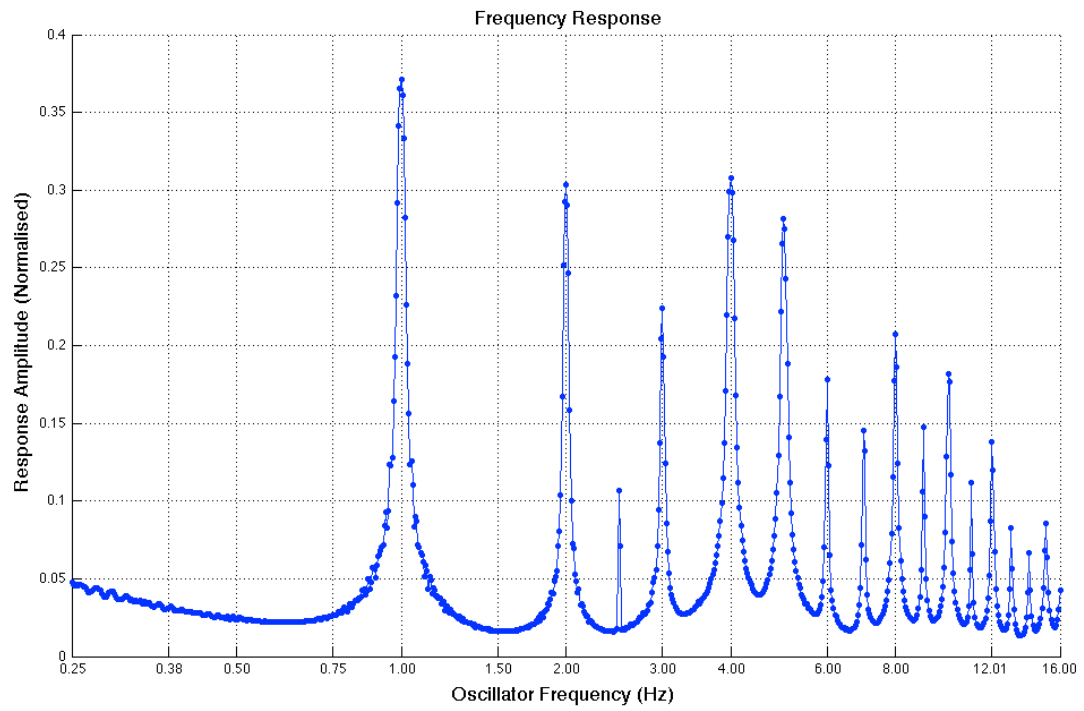
Frequency response of the canonical model in the presence of a 2:3 polyrhythm repeating once every 1000 msec.



Frequency response of the canonical model in the presence of a 2:5 polyrhythm repeating once every 1000 msec.



Frequency response of the canonical model in the presence of a 3:5 polyrhythm repeating once every 1000 msec.



Frequency response of the canonical model in the presence of a 4:5 polyrhythm repeating once every 1000 msec.

References

- Handel, S. (1984). Using Polyrythms to Study Rhythm. *Music Perception*, 1(4): 465-484.
- Handel, S., & Oshinsky J. (1981). The meter of syncopated auditory polyrythms. *Attention, Perception & Psychophysics*, 30(1): 1-9.
- Hoppensteadt, F.C., & Izhikevich, E.M. (1997). *Weakly Connected Neural Networks*, Springer.
- Iversen, J.R., Repp, B.H., & Patel, A.D. (2009). Top-Down Control of Rhythm Perception Modulates Early Auditory Responses. *Neurosciences and Music III: Disorders and Plasticity*, 1169: 58-73.
- Large, E.W. (2011). Music Tonality, Neural Resonance and Hebbian Learning. In *Proceedings of Mathematics and Computation in Music 2011*, 115-125.
- Large, E.W., Almonte F.V., & Velasco, M.J. (2010). A canonical model for gradient frequency neural networks. *Physica D: Nonlinear Phenomena*, 239(12): 905-911.
- Large, E.W. (2008). Resonating to musical rhythm: theory and experiment. In *The Psychology of Time*. Ed. S.Grondin, 189–231, Emerald, United Kingdom.
- Large, E.W. (2000a). On synchronizing movements to music. *Human Movement Science*, 19, 527-566.
- Large, E.W., & Kolen, J.F. (1994). Resonance and the perception of musical meter. *Connection science*, 6: 177.
- London, J. (2004). *Hearing in time: Psychological aspects of musical meter*. Oxford: Oxford University Press.
- Pressing, J., & Jolley-Rogers, G. (1997). Spectral properties of human cognition and skill. *Biological Cybernetics*, 76(5): 339-347.

- Pressing, J., Summers, J., & Magill, J. (1996). Cognitive Multiplicity in Polyrhythmic Pattern Performance. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5): 1127-1148.
- Repp, B. H. (2008c). Perfect phase correction in synchronization with slow auditory sequences. *Journal of Motor Behavior*, 40, 363-367.
- Repp, B. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review*, 12(6): 969-992
- Sethares, W. A. and T. W. Staley (1999). Periodicity transforms. *Signal Processing, IEEE Transactions on* 47(11): 2953-2964.
- Smith, L. M. and H. Honing (2008). Time-frequency representation of musical rhythm by continuous wavelets. *Journal of Mathematics and Music*, 2(2): 81-97.
- Snyder, J. S., & Large, E.W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research*, 24(1): 117-126.
- Snyder, J. S., & Krumhansl, C. L. (2001). Tapping to Ragtime: Cues to pulse finding. *Music Perception*, Vol. 18, No. 4 (Summer 2001), pp. 455-489
- Wilson, H.R., & Cowan, J.D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13(2), 55–80.