CONFUSED CHOIRS? A DYNAMIC GAME THEORY MODEL OF CHOIR SINGING.

Óscar Alexander Lyle Pérez, Nicholas Paul Waltz[1]
B.Sc Economics
2nd year
University College London

Explore Econ Undergraduate Research Conference
February 2020

[1] Special thanks to Prof. Martin Cripps and Prof. Dennis Kristensen for their very helpful input.

1 Introduction

Unaccompanied vocal music plays a central role in western culture but it requires excellent skill on the part of singers to achieve proper intonation. Even among professionals, the intended intonation is rarely achieved (Dai *et al.* 2015), rather, singers tend to sing marginally lower (flat) or higher (sharp). This becomes especially important in the context of a cappella choir music, where poor intonation can cumulatively result in unintended frictions and the disintegration of harmonies. Besides individual skill, experience and frequency of practice affecting pitch accuracy, singers will rely on points of reference (Mauch *et al.* 2014), which in the absence of instrumental support consist of the singers surrounding them (Alldahl 2006). This means that on occasion it may be advantageous to the overall choir quality to deviate from the intended pitch and reduce pitch variance between singers, and thus, on occasion, we observe entire choirs going sharp and flat. Based on these observations, we aim to create a game theory based model that explains behaviourally, how it is that choirs go out of tune, using the facts of biological randomness in the formation of vocal sounds, imperfect memory, and that singers cross-reference.

2 Theoretical Background

2.1 Determinants of Pitch Accuracy

It is widely accepted that perfect pitch is rarely achieved due to biological constraints in the formation of vocal sounds (Alldahl 2006). Further, skill and experience seem to improve choir quality, Dai *et al.* (2015) find a strong correlation between pitch accuracy and self-reported skill. Another area of inquiry regards the effects of imperfect memory of the practiced pitch on pitch production. When provided with a reference pitch, Mauch *et al.* (2014) find that singers will remedy their uncertainty by factoring in that reference pitch into their pitch production. This way, perturbances in the reference pitch affect intonation even after their initial occurrence.

2.2 Link to Economics

The approaches of game theory and statistical modelling used in this paper are the same underlying the modern theories of economics and repeated games in particular have vast applications, for instance in modelling the behaviour of financial markets. The fact that these seemingly specific approaches can be used effectively in the context of choir intonation, a field of inquiry perhaps as distant from economics as they come, goes to show the breadth of applications of economic tools.

3 Modelling

We will model choir intonation over time, by first establishing the singers as utility maximising, or rather loss minimising, entities in the face of imperfect pitch memory. We then introduce individual and biological randomness, add a time structure to memory itself, and finally construct a determined system based on backward-looking behaviour.

3.1 Definitions

In the following, a "choir" are N homogenous singers $i = 1,2,3, \ldots, N$ that have to sing the same melody of T notes $n_1^*, n_2^*, \ldots, n_T^*$. In the context of deficiencies in memory and biological randomness, we will further define p_{it} as singer i's *targeted* pitch at note t and s_{it} as the *realisation* of the t^{th} note. Lastly, by *drift* or *pitch error* we mean the deviation $e_{it} = s_{it} - n_t^*$.

3.1a *Metrics*

In our analysis, we make use of the measure of "cent" used in the file type MIDI[2]. We relate the fundamental frequency f_0 to musical pitch p in cents in the following way

$$p = 6900 + 1200 \log_2 \frac{f_0}{440}$$

This way, a drift of +100 cents means going sharp by one semitone. In our analysis, we further use the mean of the squared pitch errors as a measure of dissonance in the choir

$$MSPE_t = \frac{1}{N} \sum_{i=1}^{N} e_{it}^2$$

3.2 Pitch Production

3.2a *Pitch Memory and Preferences*

The main reason that choirs go out of tune is that individual singers have imperfect pitch memory. This way, the singers' targeted pitch will be affected, when the singer is provided with a reference pitch (Mauch *et al.* 2014). And in the context of a cappella choir singing, the singers' reference pitches are the impression they gain from the other singers. In particular, reference pitch r_{it} can be said to be singer i's expectation of what the others' average pitch realisation will be at note t, so

$$r_{it} = \frac{1}{N-1} \sum_{j \neq i} s_{jt}^E$$

Here, we adopt the concept of pitch memory introduced by Mauch *et al.*, using the measure $\mu \in [0,1]$, where $\mu = 0$ means no memory, or full reliance on the reference pitch, while $\mu = 1$ means absolute or perfect memory and so no reliance on a reference pitch whatsoever.

[2] See Vurma and Ross (2006), and White (1999) for more detail.

Using this, singers can be said to target both the precise pitch, and the reference pitch simultaneously, weighted by their pitch memory and will therefore derive disutility from deviation from the actual pitch over their active memory, and from the reference pitch, where they are uncertain. As it is reasonable to assume that singers also care about the choir sounding good after spending months practicing a song, we add the disutility singers experience from the realised total dissonance among the other singers as a further quantification. Hence, we define the loss of singer i at the t^{th} note as

$$L_{it} = \mu(p_{it} - n_t^*)^2 + (1-\mu)(p_{it} - r_{it})^2 + \sum_{j \neq i}\left(s_{jt} - n_t^*\right)^2$$

Given the particular loss function above, singer i's best response is then to target (see Appendix 6.1)

$$p_{it} = \mu n_t^* + (1-\mu) r_{it},$$

which directly replicates the model by Mauch *et al.* for our case.

3.2b *Realised Pitch*

Due to the biological nature of vocal singing, however, we have to make a distinction between the pitch, singers target and the pitch realisation that results. The realised pitch is to some extent inherently random. This is also, why experience and skill matter: among other things, training increases accuracy by reducing random errors. For the pitch realisation s_{it}, we therefore add to the targeted pitch a Gaussian pitch error $\epsilon_{it} \sim N(0, \sigma_i^2)$, which captures these biological imperfections, as well as other individual characteristics, such as skill[3], and will likely also increase with the complexity of the song itself. So,

$$s_{it} = p_{it} + \epsilon_{it}$$

Due to the cross-referencing described above, individual perturbances will be picked up by the other singers thus are the reason that we observe a cappella choirs going out of tune in the real world.

3.3 The Time Structure of Pitch Memory

A cappella choirs are typically given an initial note or chord on an instrument, typically on a piano, organ or tuning fork. This essentially provides singers with an exact reference pitch, i.e.

$$r_{i1} = n_1^*, \text{ so } p_{i1} = n_t^*.$$

Over time, the memory of that perfect initial reference pitch will diminish and for their targeted pitch, singers have to increasingly rely on their individual sense of musical temperament, i.e. their individual pitch memory, which is improved mainly through practice but will also affected by other individual characteristics, such as skill and experience.

[3] Mauch *et al.* find mean absolute deviations of 0.6 semitones for self-declared amateur singers and 0.3 semitones for self-declared professional singers.

The initial chord or note essentially decreases reliance on a reference pitch, as the reference itself now contains an accurate sense of the actual note n_t^* that decreases over time. For this reason, we can say that it is pitch memory itself that has a time structure. In particular, $\mu_t = 1$, i.e. perfect pitch memory, at $t = 1$ and eventually $\mu_t = \bar{\mu}$, the individual sense of musical temperament, when $t \gg 1$. We model this as exponential decay, with 1 as the intercept and $\bar{\mu}$ as the asymptote, using $\alpha > 0$ as a measure of the rate of memory decline.

$$\mu_t = (1 - \bar{\mu})e^{-\alpha(t-1)} + \bar{\mu}$$

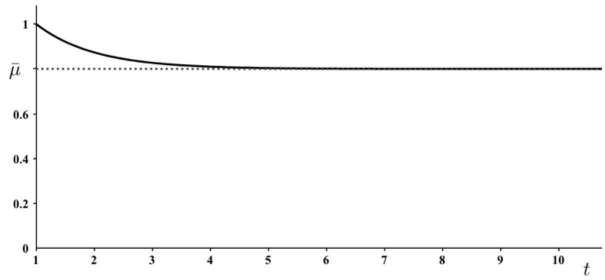

Figure 1: An example of the time path of pitch memory with $\bar{\mu} = 0.8$ and $\alpha = 1$.

3.4 The Repeated Game

The choir game will therefore have the following structure. At the first note, singers have perfect memory ($\mu_1 = 1$), so their targeted pitch is $p_{i1} = n_1^*$ and we expect all drift to come from random disturbances due to individual and biological characteristics, i.e.

$$e_{i1} = \epsilon_{i1} \sim N(0, \sigma_i^2).$$

For the notes following the first, $\mu_t < 1$, so singers will in part rely on their impression of what the remaining singers' realisations are. Since singers will certainly not be affected by future realisations, singers will only factor in present and previous realisations, and therefore are at least in part backward looking. Because the most recent and informative reference that singer i has, is the perception of the others' drifts one note prior, $e_{j,t-1}$, i's estimation of j's realisation s_{jt}^E, can be said to factor in this exact deviation plus some bias b_{ijt} that captures other factors influencing pitch estimation, such as interval size in the melody (e.g. if i has a tendency to underestimate intervals), perceived relative skill (e.g. i is an amateur and knows that j is professional), distance between the singers, and ear training (ability to recognise pitch)[4].

$$s_{jt}^E = n_t^* + e_{j,t-1} + b_{ijt}, j \neq i$$

Or in terms of the reference pitch (see Appendix 6.2),

[4] According to Alldahl, singers are more likely to go flat than sharp. This tendency could be modelled by having a negative bias.

$$r_{it} = n_t^* + \bar{e}_{-i,t-1} + b_{it},\ b_{it} = \frac{1}{N-1}\sum_{j\neq i} b_{ijt}$$

Where $\bar{e}_{-i,t-1}$ is the average drift of the others realised at the previous note. For our model, by assumption of homogeneity and for simplicity, we say $b_{ijt} = 0$, without affecting the qualitative results, so ceteris paribus, singers are only influenced by the average drift of the others they perceived one note prior. From this, we get the full expression for singer i's pitch realisation at note t (see Appendix 6.2).

$$s_{it} = n_t^* + (1-\mu_t)\bar{e}_{-i,t-1} + \epsilon_{it}$$

And for the realised drift $e_{it} = (1-\mu_t)\bar{e}_{-i,t-1} + \epsilon_{it}$.

3.5 Results

3.5a *Qualitative Results*

Given the particular disturbances ϵ_{it}, we can construct a determined system, which takes the form (see Appendix 6.3)

$$\boldsymbol{e_t} = \sum_{j=1}^{t-1} \frac{A^j \epsilon_{t-j}}{(N-1)^j} \prod_{k=0}^{j-1}(1-\mu_{t-k}) + \boldsymbol{\epsilon_t}, \boldsymbol{e_1} = \boldsymbol{\epsilon_1}$$

With $\boldsymbol{e_t} = \begin{bmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{Nt} \end{bmatrix}$, $\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 1 & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & 1 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix}$ and $\boldsymbol{\epsilon_t} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{Nt} \end{bmatrix}$.

From this we can derive the variance of individual drift over time (see Appendix 6.3), which, in combination with the fact that $E[e_{it}] = 0$, we can use to construct confidence intervals, which attach a probability to drift being contained in a certain interval about 0 (see Figure 2).

$$Var(e_{it}) = \sigma^2\left(1 + \sum_{j=1}^{t-1} \frac{1}{(N-1)^j} \prod_{k=0}^{j-1}(1-\mu_{t-k})^2\right)$$

We find that the drift confidence intervals get tighter with skill (through σ_i^2), practice (through the sense of musical temperament $\bar{\mu}$), and with the number of singers (see Appendix 6.4). We also find that the confidence intervals for drift only increase a little initially and eventually stabilise. The first two results, we would intuitively expect. The fact that more singers improve the variance is a consequence of singers factoring in the average pitch drift of the others, which, as numbers increase, is more likely close to 0. This assumes that all singers equally affect individual drift, however. The last result, though not directly intuitive, is also representative of the real world. If the variance were to increase without bounds, this would mean that over time, choir singing becomes very volatile, something that we do not observe. Rather, individual singers will always in part have the tendency to be on pitch as this is what they practiced and so, through cross-referencing, choir quality will stabilise.

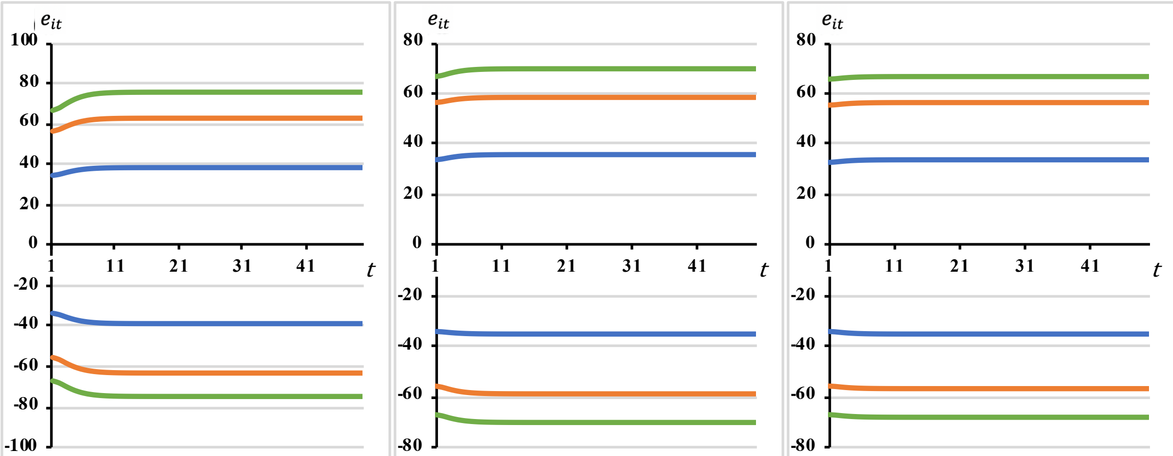

Figure 2: The 68% (blue), 95% (orange) and 99% (green) confidence intervals of individual drift over time for different parameters (left: $\bar{\mu} = 0.2$, $N = 4$; middle: $\bar{\mu} = 0.8$, $N = 4$; right: $\bar{\mu} = 0.2$, $N = 20$)

3.5b *Using the Model, an Example: Choir Goes Sharp at the 5th Note*

Here, we show how the model can be applied. We consider the case of $N = 4$ amateur singers with $\sigma_i = 73$[5] that have to sing the melody $n^* = (n_1^*, n_2^*, \dots, n_T^*)$. Further, we say that we are modelling choir quality during a practice session before performance, where the individual sense of temperament is low, at $\bar{\mu} = 0.5$.

Say that, at note $t = 5$, two singers go sharp by 70 cents each (0.7 semitones). With $\sigma_i = 73$, the probability of a singer drifting by more than 70 cents per singer is $\sim 17\%$, so approximately once every five notes.

$t = 5$	$e_{1,5} = 0$	$e_{2,5} = +70$	$e_{3,5} = +70$	$e_{4,5} = 0$	$MSPE = 2450$

Then, using $E_{it}[e_{it}] = E\left[e_{it} \middle| \bar{e}_{-i,t-1}\right] = (1 - \mu_t)\bar{e}_{-i,t-1}$, i.e. assuming no further disturbances, and assuming $\alpha = 1$, we get

Note (t)	$E_{1t}[e_{1t}]$	$E_{2t}[e_{2t}]$	$E_{3t}[e_{3t}]$	$E_{4t}[e_{4t}]$	$MSPE$
6	23.1761146	11.5880573	11.5880573	23.1761146	337.15031
7	7.74314954	9.67273127	9.67273127	7.74314954	76.6550475
8	4.50911842	4.18762935	4.18762935	4.50911842	18.9432993
9	2.14782038	2.20139528	2.20139528	2.14782038	4.72889055
10	1.09160406	1.08267532	1.08267532	1.09160406	1.18195516
…	…	…	…	…	…

[5] Based on the finding by Mauch *et al.* that for amateur singers the mean absolute deviation d is 60 cents, and for normal distributions, $d = \sqrt{2\sigma^2/\pi}$.

Since deviations of as low as one cent can be perceived by humans[6], ceteris paribus the initial disturbance results in the choir going audibly sharp for the following periods. In the real world, we indeed observe choirs going sharp like this and our model now provides a way of behaviourally explaining, how this might come about.

4 Conclusion

We have created a simple model of a cappella choir singing for the case of backward-looking homogenous singers with the same melody. For this, we have made use of methods usually applied in an economic context to analyse, why choirs go out of tune. We think, this goes to show that these methods, because they were initially constructed to describe basic human behaviour quantitively, are indeed descriptive outside of the economic context and perhaps fundamental in some way. For future work, a more compelling model of choir singing will relax our assumptions and factor in heterogeneities, most notably the effects of different registers (Soprano, Alt, Tenor, Bass) on pitch production. Further areas of research include the effect of just-noticable frequency differences, harmonic tension, and continuous time.

Word count: 2097

6 Appendix

6.1 Best Response

The loss function is

$$L_{it} = \mu(p_{it} - n_t^*)^2 + (1-\mu)(p_{it} - r_{it})^2 + \sum_{j \neq i} \left(s_{jt} - n_t^*\right)^2$$

This is minimised, where

$$\frac{\partial L_{it}}{\partial p_{it}} = 0.$$

So, where

$$\frac{\partial L_{it}}{\partial p_{it}} = 2\mu(p_{it} - n_t^*) + 2(1-\mu)(p_{it} - r_{it}) = 0.$$

Rearranging gives the best response

[6] See Pierce (1983) for a breakdown of just-noticable differences given different signal strengths and frequency ranges.

$p_{it} = \mu n_t^* + (1-\mu) r_{it}$.

6.2 Reference Pitch and Pitch Realisation under Backward Looking Behaviour

Using $s_{jt}^E = n_t^* + e_{j,t-1} + b_{ijt}$ and $r_{it} = \frac{1}{N-1}\sum_{j\neq i} s_{jt}^E$,

$r_{it} = \frac{1}{N-1}\sum_{j\neq i} n_t^* + e_{j,t-1} + b_{ijt} = n_t^* + \frac{1}{N-1}\sum_{j\neq i} e_{j,t-1} + b_{it}$
$= n_t^* + \bar{e}_{-i,t-1} + b_{it}$, with $b_{it} = \frac{1}{N-1}\sum_{j\neq i} b_{ijt}$

Since, $s_{it} = \mu n_t^* + (1-\mu) r_{it} + \epsilon_{it}$, we get $s_{it} = n_t^* + (1-\mu)\bar{e}_{-i,t-1} + \epsilon_{it}$ and
$e_{it} = s_{it} - n_t^* = (1-\mu)\bar{e}_{-i,t-1} + \epsilon_{it}$

6.3 Dynamic System

By assumption, $s_{it} = \mu n^* + \frac{(1-\mu_t)}{N-1}\sum_{j\neq i}(n_t^* + e_{j,t-1}) + \epsilon_{it}$, so

$$e_{it} = s_{it} - n_t^* = \frac{(1-\mu_t)}{N-1}\sum_{j\neq i} e_{j,t-1} + \epsilon_{it}$$

With $\mu_t = (1-\bar{\mu})e^{-\alpha(t-1)} + \bar{\mu}$.

The system therefore takes the form

$$\boldsymbol{e_t} = \frac{1-\mu_t}{N-1}\boldsymbol{A}\boldsymbol{e_{t-1}} + \boldsymbol{\epsilon_t}$$

With $\boldsymbol{e_t} = \begin{bmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{Nt} \end{bmatrix}$, $\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 1 & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & 1 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix}$ and $\boldsymbol{\epsilon_t} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{Nt} \end{bmatrix}$

Given $\boldsymbol{e_1} = \boldsymbol{\epsilon_1}$, as $\mu_1 = 1$, for $t > 1$, we get

$$\boldsymbol{e_t} = \sum_{j=1}^{t-1} \frac{\boldsymbol{A}^j \boldsymbol{\epsilon_{t-j}}}{(N-1)^j} \prod_{k=0}^{j-1}(1-\mu_{t-k}) + \boldsymbol{\epsilon_t}$$

Then, given $\epsilon_{it} \sim N(0, \sigma_i^2)$,

$$E[\boldsymbol{e_t}] = 0$$

The Gaussian pitch errors are independent, so for $t > 1$

$$Var(\boldsymbol{e_t}) = \sum_{j=1}^{t-1} \frac{1}{(N-1)^{2j}} \prod_{k=0}^{j-1} (1-\mu_{t-k})^2 \, Var\left(\boldsymbol{A}^j \boldsymbol{\epsilon_{t-j}}\right) + Var(\boldsymbol{\epsilon_t})$$

Lastly by assumption of homogeneity, $Var(\epsilon_{it}) = \sigma_i^2 = \sigma^2$, so we get

$Var(\boldsymbol{e_1}) = \boldsymbol{\sigma^2}$

$Var(\boldsymbol{e_t}) = \boldsymbol{\sigma^2} \left(1 + \sum_{j=1}^{t-1} \frac{1}{(N-1)^j} \prod_{k=0}^{j-1} (1-\mu_{t-k})^2\right), t > 1$, where $\boldsymbol{\sigma^2} = \begin{bmatrix} \sigma^2 \\ \sigma^2 \\ \vdots \\ \sigma^2 \end{bmatrix}$

Or, $Var(e_{it}) = \sigma^2 (1 + \sum_{j=1}^{t-1} \frac{1}{(N-1)^j} \prod_{k=0}^{j-1} (1-\mu_{t-k})^2)$

6.4 Effect of Changes in Parameters on Variance

Variance decreases with increases in the number of singers in periods following the first:

$\frac{\partial Var(e_{it})}{\partial N} = 0, \ t = 1$

$\frac{\partial Var(e_{it})}{\partial N} = -\sigma^2 \sum_{j=1}^{t-1} \frac{1}{(N-1)^{j-1}} \prod_{k=0}^{j-1} (1-\mu_{t-k})^2 < 0, t > 1$

Variance decreases with an increased sense of musical temperament (practice) in periods following the first:

$\frac{\partial Var(e_{it})}{\partial \bar{\mu}} = 0, \ t = 1$

Using $(1-\mu_t) = (1-\bar{\mu})(1-e^{-\alpha(t-1)})$,

$\frac{\partial Var(e_{it})}{\partial \bar{\mu}} = -\sigma^2 \sum_{j=1}^{t-1} \frac{2j(1-\bar{\mu})^{2j-1}}{(N-1)^j} \prod_{k=0}^{j-1} \left(1-e^{-\alpha(t-k-1)}\right)^2 < 0, t > 1$, as $1-\bar{\mu} \geq 0$

Variance decreases with increased skill ($\sigma \downarrow$) in all periods:

$\frac{\partial Var(e_{it})}{-\partial \sigma^2} = -1, t = 1$

$\frac{\partial Var(e_{it})}{-\partial \sigma^2} = -\left(1 + \sum_{j=1}^{t-1} \frac{1}{(N-1)^j} \prod_{k=0}^{j-1} (1-\mu_{t-k})^2\right) < 0, t > 1$

7 List of References

1. Alldahl, P.-G. (2006), *Choral Intonation*, Stockholm, SE: Gehrmans Musikförlag.
2. Dai, J., Mauch, M., Dixon, S. (2015), *Analysis of intonation trajectories in solo singing*, London, UK: Centre for Digital Music. Queen Mary University of London.
3. Weiß, S., Schlecht, S., Rosenzweig, S., Müller, M. (2019), *Towards Measuring Intonation Quality of Choir Recordings: A Case Study on Bruckner's Locus Iste*, Erlangen, DE: International Audio Laboratories.
4. Mauch, M., Frieler, K., Dixon, S. (2014), *Intonation in Unaccompanied Singing: Accuracy, Drift and a Model of Reference Pitch Memory*, London, UK: Centre for Digital Music. Queen Mary University of London.
5. Pierce, John R. (1983), *The Science of Musical Sound / John R. Pierce*. New York: Scientific American. Print. Scientific American Library Ser.
6. Vurma, A. and Ross, J. (2006), *Production and perception of musical intervals*, Music Percept. 23, 331–344.
7. White, P. (1999), *Basic MIDI,* London: SMT