9th International Conference on Music Perception and Cognition
Alma Mater Studiorum University of Bologna, August 22-26 2006

A comparison between the temporal and pattern approach to virtual pitch applied to the root detection of chords

Ludger Hofmann-Engl
Dept. of music, The Link Schools
London, England
email hofmann-engl@chameleongroup.org.uk

ABSTRACT

This paper attempts to address the issue of virtual pitch from an empirical point of view comparing two specific virtual pitch models. One of these models belongs to the class of the temporal approach models and the other model belongs to the class of the pattern approach models. The choice of these specific models was intentional as they can be considered to be the most sophisticated models within their class.

A preliminary experiment together with an experiment proper produced data which are in support for the pattern model. Particularly in cases were uncommon chords were involved, the temporal model failed to make any predictions. Additionally, it was shown that familiarity and context have an influence on root detection and that significant order effects can be observed.

Keywords

Virtual Pitch, temporal approach, pattern approach, familiarity, order effect, common and uncommon chords, root detection.

1 INTRODUCTION

For the last 3 decades (since Terhardt's introduction to the pattern approach to virtual pitch in 1976, 1977, 1979), the debate has been raging on whether the temporal approach to virtual pitch represents the holy grail or perhaps the pattern approach. The fact, that Moore (1997) pronounced the pattern approach more or less dead, did not help to continue the debate on unbiased grounds and Terhardt appears to have turned into a rather disillusioned reclusive. Interestingly, however, none of the multitude of virtual pitch models have been put the test empirically with the exception of Hofmann-Engl's model (1990). Testing root preferences on a sample of almost 100 participants, Hofmann-Engl produced good support for his pattern model. Additionally, Hofmann-Engl (1999) demonstrated that his model is a useful compositional tool and in 2004 he showed that the model can be a powerful analytical tool in the context of contemporary classical music.

However, a major shortcoming of the experiments performed by Hofmann-Engl (1990) must be seen in the fact that Hofmann-Engl did not compare the pattern model with a temporal model. This is, no evidence was produced that, while the pattern model appeared to be generally reliable, a temporal model would perform similarly well or perhaps even better.

The reason why the author chose Meddis's and Hewitt's model as the representative one for the temporal approach is due to the fact that not only can this model be regarded as the most sophisticated model within its class, but that the main part of it it exists as a software package called AMS (Auditory Model Simulator) at:

http://www.essex.ac.uk/psychology/hearinglab/dsam/dloads_main.htm,

which allows the actual input of auditory signals. According to Meddis and Hewitt (1991), the output of this simulator in form of a summery autocorrelation function based on the work of Licklieder (1959), can be regarded as the foundation of a root detector. While Meddis and Hewitt (1991) interpreted the peak of the summery ACF to represent the virtual pitch of the given stimulus, a different methodology was implemented for this experiment. This methodology is explained below.

Hofmann-Engl's model too is available as a software package in form of an applet called *Harmony Anayzer3.0* at:

http://www.chameleongroup.org.uk/software/piano.html

whereby tones are inputed via a keyboard and a list of virtual pitches is calculated inside a window. Additionally, the first five best roots can be made to sound together with the chord that has been inputed.

2 ROOT DETECTION

2.1. The Temporal Model

The model as introduced by Meddis and Hewitt (1991) is an attempt to describe the auditory process in form of a detailed physiological simulation including factors such as basilar membrane filtering, neuro transmission, refraction time of fibers, ect. A detailed description of the model can be found in Meddis and Hewitt (1991). The model is sophisticated and its application in form of the AMS software package allows for the input of recorded wav files. For a number of reasons the method of picking the peak of the ACF to represent the the root (virtual pitch) of the input signal was abandoned because a) in several cases the peak differed fundamentally when compared to the results during the initial test period (compare Hofmann-Engl , 2004) b) several stimuli (midi single tones) produced peaks in the distance of one octave at times with all peaks of the same strength and in case of the violin sound with almost linearly decreasing intensity for increasing frequencies and c) not every chord produced a single clear peak nor comparable peaks in octave distance.

Thus, initial idea was to compare the ACF of the test stimulus with the ACF of a sinusoidal tone and to vary the frequency of the sinusoidal tone until a minimal Euclidean distance between the ACFs was computed. However, the results were not satisfying especially for test stimuli which produced peaks at octave intervals. Further testing showed that comparison stimuli consisting of 5 harmonic partials produced far better results, and hence all test stimuli were compared to control stimuli consisting of 5 harmonics (adding more harmonic did not seem to make any difference).

Finally, during a number of preliminary test trials, it appeared that the Euclidean distance produced less reliable results than did an exponential distance. Various variables were used, until it was decided that the following distance appeared to be most consistent:

$$D(p)=\sum_{i=1}^{n} e^{-10[A(ACF_t)_i - A(ACF_c(p))_i]^2} \qquad (1)$$

where $D(p)$ is the distance between the ACF_t of the test stimulus and the ACF_c of the control stimulus at the pitch *p*, *n* is 1/sampling rate, $A(ACF_t)_i$ is the amplitude of the summery autocorrelation function of the test stimulus at the frequency *i* and $A(ACF_c(p))_i$ is the amplitude of the summery autocorrelation function of the comparison stimulus at the pitch *p*.

The minimal distance $D(p)_{min}$ was used (compare Appendix) to detect the roots to the chords (according to the temporal model) used during the preliminary experiment and the experiment proper.

2.2. The pattern model

The pattern model according to Hofmann-Engl was used in it's original form as deviced in 1990. The formula for the root detection is the following one:

$$V(t)=\sum_{i=1}^{n} \frac{w_s(s_i)w_p(s_i)}{n} \qquad (2)$$

where $V(t)$ is the strength of the virtual tone *t*, $w_s(s_i)$ the spectral weight of the *i*th subharmonic of the chord, $w_p(s_i)$ the weight of the *i*th subharmonic according to the position of the tone within the chord, *n* the number of tones the chord consists of and the constant $c = 6\ Hh$

A detailed description of the calculations can be found in Hofmann-Engl (1990, 1999, 2004).

3. PRELUMINARY EXPERIMENT

6 composers from the Royal Academy of Music took part during a preliminary experiment.

The concept of the experiment was the following. Roots were calculated according to both virtual pitch models for some common and some uncommon chords. For uncommon chords composers were asked whether they preferred the root according to the pattern model or according to the temporal model. The experiment was programmed in java and the participants were given the option to either prefer root(pattern), root(temporal), '*equal*' or '*not sure*'. All stimuli were played in order A – B and B – A. For common chords (such as major/minor chords) not the best root was compared but arbitrarily the 5^{th} best according to both models. Participants heard the chord followed and overlapping by the bass note (root). The bass note (*c*) was the same throughout the experiment.

Three main effects were observed: A) The bass note *c* was too high and did not generate the sense of a root, b) predictions on common chords did neither make sense according to the pattern model nor to the temporal model and c) the root of none common chords were predicted according to the pattern model and not according to the temporal model. Interestingly, the comparison of the chord c, d ,e with root c according to the pattern model was preferred by all participants in both orders in contrast to the chord a, b ,c# which predicted c to be the root acording to the temporal model. It was decided to repeat the experiment with a changed bass note (a) on 14 amateur and semi-and professional musicians.

4. EXPERIMENT PROPER

The preliminary experiment had indicated that somehow the predictions for common chords seem not to work even after avoiding familiarity effects by comparing the 5^{th} best root according to the temporal pattern with the 5^{th} best root according to the pattern model. The experiment proper involved the same stimuli as did the preliminary experiment except that the bass note was changed from the note (c) to the note *a*.

4.1.Method

Participants

A sample of 14 amateur, semi and professional musicians participated during the experiment proper. 6 of the participants had been trained in classical piano, 3 in Jazz piano, 2 in Jazz Saxophone, 2 in Rock guitar and one in Rock (Ballads) vocals. 7 of the participants were female and 7 male (mean age ca. 30).

Equipment

A laptop (acer, TravelMate 4050) running under windows xp was programmed in Java and the audio output was presented to the participants via headphones (Philips, SBC HS900) at moderate level of loudness . The audio signals were midi piano signals. An external mouse (Genius GM 03021 U/A) was used.

Stimuli

The stimuli were presented randomized in both orders *a – b* and *b – a* (all in all 28 stimuli). Each stimulus consisted of two events: Events consisted of a chord (start: 0 midi units – end: 25 midi units) and a bass note A (start: 8 midi units – end: 23 midi units). This is, chord and bass note overlapped by 14 midi units. The reason why the bass note was cut off at 23 units was based on the fact that due to reverberation of the midi signal the bass note faded together with the chord at 25 units. The chords were the following:

Table 1. List of Experimental Stimuli

Trial/Chord	*A*	*B*	*V(a)*	*D(a)*
1 (5^{th})	*c#, f, g#*	*f#,a#, c#1*	*2.13 Hh*	*1.03*
2 (5^{th})	*g, a#, d1*	*d, f, a*	*1.50 Hh*	*1.15*
3 (5^{th})	*g#, c#1, f1*	*d, g, b*	*1.87 Hh*	*1.7*
4 (5^{th})	*a, d, f*	*b, e1, g1*	*2.00 Hh*	*3.35*
5 (5^{th})	*b, d1, g1,*	*f#, a' d1*	*1.98 Hh*	*1.41*
6 (5^{th})	*c#, f, a#*	*d, f#, b*	*1.78 Hh*	*0.64*
7 (1^{st})	*g, g#, c#1*	*e, f, a#*	*2.96 Hh*	*1.94*
8 (1^{st})	*g, g#, a*	*e, f, f#*	*3.09 Hh*	*1.94*
9 (1^{st})	*c#, g, g#*	*a, d#, e*	*3.19 Hh*	*3.12*
10 (1^{st})	*g, g#, b, c#1*	*e, f, g#, a#*	*2.6 Hh*	*1.72*
11 (1^{st}))	*c#, g, b, e1*	*e, a#, d1, g1*	*3.34 Hh*	*2.02*
12 (1^{st})	*a, b, c#1*	*f#, g#, a#*	*3.81 Hh*	*0.43*
13 (1^{st})	*c#, e, g, a#*	*e, g, a#, c#*	*3.01 Hh*	*2.92*
14 (1^{st})	*a, c1, e1, g1, b1*	*d, f, a, c1, e1*	*2.62 Hh*	*1.21*

As observed by Hofmann-Engl (2004) it appears to be difficult to test familiar chords with unfamiliar chords or chords with different degree of familiarity. Hence the author decided not to test familiar chords against the strongest virtual pitch but against the fifth strongest candidate. We take chord A from trial 1 as an example.

According to the pattern model, we obtain the following data for the chord *c#*, *f*, *g#*:

Table 2. Virtual Pitch Data for the chord c#, f, g# according to Harmonyanalyzer 3.0

Virtuality of: c# = 4.38 Hh	Virtuality of: g# = 1.15 Hh
Virtuality of: f# = 3.02 Hh	Virtuality of: b = 1.11 Hh
Virtuality of: d# = 2.29 Hh	Virtuality of: g = 1.06 Hh
Virtuality of: a# = 2.24 Hh	Virtuality of: e = 1.03 Hh
Virtuality of: a = 2.13 Hh	Virtuality of: d = 0.61 Hh
Virtuality of: f = 1.41 Hh	

Here, the fifth best match is (a). However, according to the temporal model the fifths best match to the bass note a is the chord *f#*, *a#*, *c*. Here,the hypothesis is that should the pattern model be correct, a preference for chord A ought to be expected and if the temporal model was to be correct a preference for chord B ought to be detected.

For uncommon chords the best match (root) was computated according to both models. Considering stimulus 8, we find that the best match for a chord consisting of two minor seconds is the chord *g*, *g#*, *a* in relation to the bass note a according to the pattern model and the chord *e*, *f*, *f#* according to the temporal model.

Procedure

Participants were instructed through an example on the piano (c-major with bass not (c) compared with f#-major with bass note (c)).

During the experiment, participants had to enter their names via a TextField. Once they had entered their name a start button appeared. After pressing *Start* candidates were informed on how many trials they would have to listen to. This information was updated with each completed trial. At the beginning of each trial a box appeared labeled *sound a*. Once the participant had listened to *sound a* a second box appeared labeled *sound b*. Once *sound b* was heard a headline appeared below the two sound boxes saying: Best match, with four active buttons below labeled: *A – B – equal – not sure*. The results for each participants were stored as a data file after the completion of the 28 trials.

4.2.Results

4.2.1.Order Effects

All trials were played in the order *a – b* and *b – a*. It was expected that order ought not be of statistical significance. However, when comparing the answers given as A when played in order *a – b* with the answers A given when played in order *b – a,* we obtain a mean for order *a – b* of 6.8 *A*s and 4.6 *A*s for the order *b – a*. The ANOVA confirms that this difference is significant with $F(1, 14) = 6.3$ and $p < 0.02$.

Order effects are particularly compelling in trials 11 and 14. Particularly trial 14 is interesting, in order *b – a* trial 14 delivers a significant preference for *Bs* while there is a non significant preference for *A*s. While four participants selected *A* in trial 14 in order *a – b*, only one single participant preferred *A* in the order *a – b.*

4.2.2 Root/5th best match detection

As each trial was played in the order *a – b* and *b – a.* Each answer *a* fetched the value 1, each answer *b* the value -1 and each answer *equal* and *not sure*, the value 0. The results are listed below (the overall rating is simply obtained by adding the numbers):

Table 3. List of all results from all trials combined in order a – b and b – a (with s for significant, ns not significant and as approaching significance

Trial	A	B	Equal/not sure	Overall rating
1 (5th)	**11 * 1**	**7 * -1**	10 * 0	4 ns
2 (5th)	5 * 1	15 * -1	8 * 0	-10 s
3 (5th)	11 * 1	9 * -1	8 * 0	2 ns
4 (5th)	19 * 1	4 * -1	5 * 0	15 s
5 (5th)	7 * 1	15 * -1	6 * 0	-8 ns
6 (5th)	15 * 1	8 * -1	5 * 0	7 ns
7 (1st)	13 * 1	5 * -1	10 * 0	8 as
8 (1st)	9 * 1	7 * -1	12 * 0	2 ns
9 (1st)	14 * 1	9 * -1	5 * 0	5 ns
10 (1st)	12 * 1	7 * -1	9 * 0	5 ns
11 (1st)	15 * 1	8 * -1	5 * 0	7 ns
12 (1st)	15 * 1	4 * -1	9 * 0	11 s
13 (1st)	9 * 1	12 * -1	7 * 0	-3 ns
14 (1st)	5 * 1	17* -1	6 * 0	-11 s

The result indicates that all uncommon chords (7 to 12) are correctly predicted by the pattern model. Four of the chords based on a major minor harmony appear to have been predicted correctly by the temporal model and four by the pattern model. Comparing the answers as given for *A* and *B* (in both orders) statistically, produces a difference (mean *A*s is: 11.7 and *B*s is: 8.9) approaching significance in the favor of *As* ($F(1, 14) = 3.7$ and $p < 0.07$.

However, comparing only the answers given in the order *a – b* (*A*s: 6, 3, 5, 11, 4, 8, 7, 5, 9, 9, 9, 8, 7, 4 and *B*s: 3, 8, 4, 2, 7, 3, 1, 3, 2, 2, 4, 2, 3, 5) produces a highly significant difference with $F(1, 14) = 15.9$ and $p < 0.0005$. Considering only the results in order *a – b*, we obtain significances for trial 4, 7, 9 and 10 with trial 5 and 12 approaching significance. Considering order *b – a* only trial 13 obtains significance.

4.3Discussion

In order to evaluate the results, it might be useful to consider the trials 7 to 12 first.

4.3.1.Uncommon Chords

Although single trials tend to produce no significant difference when comparing the counts of *A*s with the counts of *B*s, by comparison of the pooled data we obtain $\chi^2 = 6.1$ and $p < 0.02$, hence the As appear significantly more often than the *B*s indicating that the pattern model is a significantly better root predictor than the temporal model.

Particularly interesting is trial 12 (whole tone chord) as the difference between the data sets is highly significant and in in favor of the pattern model. Considering that trials 7, 9 and 10 produced significant support for the pattern mode in order *a – b*, it can be concluded with good certainty that the pattern model is not only superior to the temporal model, but that it appears all in all to be a reliable root detector for uncommon chords.

4.3.2 Common simple Chords

Simple common chords (major, minor triads including inversions) in trials 1 to 6 appear to have produced less clear data. Four of the roots had been predicted correctly by the pattern model and 2 by the temporal model. As much as the pattern model appears to be superior, a χ^2 test shows that with $r^2 = 0.8$ this difference is not significant.

Interestingly, a mistake in design of the experiment holds the key to understanding the situation and this mistake occurred in trial 4. It is true, that a is the 5[th] best match to the chord *a*, *d* , *f* with $V(a) = 2.00$ *Hh*. However, unnoticed by the author, the comparison chord b, e_1, g_1 produces $V(a) = 3.35$ *Hh* with the root (a) being the second best match to the chord. This means, that both the pattern and the temporal model are actually in favor of the chord b, e_1, g_1 and yet the participants of the experiment clearly "voted" against this chord. So what went wrong?

As mentioned before, the reason why the 5[th] harmonic was chosen, had been made to avoid familiarity judgments. However, the d-minor chord with the a in the base is actually nothing else but the second inversion of the d-minor chord and hence highly familiar to a listener brought up in a Western cultural environment. That the d-minor chord would be heard as being more appropriate would be further supported by the fact that the entire experiment makes use of the bass note a and hence the d-minor chord will be judged as being more appropriate because it represents the sub-dominant to a-minor. Hence, we can conclude that strong familiarity would outweigh the root as predicted by both the pattern and the temporal model.

Trial 2 adds support to this claim where again a d-minor chord was judged as being more appropriate despite it being a worse match than its comparison chord. This means, in order to test common chords in future projects more effort will have to be made in order to eliminate familiarity.

4.3.3. Common Complex Chords

Trial 14 is an interesting trial. While, one chord is the d-minor 7[th] 9[th] chord the other chord is the a-minor 7[th] 9[th] chord. Surely, one would expect the participants to prefer the a-minor 7[th] 9[th] but they did not and this with a significant difference. The author wishes to add the comment that this preference is odd whether the pattern or the temporal model are correct or not. It simply clashes with music theory except if we consider the context that all chords were played on the background of the bass note a and hence an a-major/minor atmosphere had been created. Now, the d minor 7[th] 9[th] does belong to a-minor but the a minor 7[th] 9[th] does not and hence the context outweighs the prediction according to the pattern model as well as the prediction according to traditional harmony theory.

Finally, trial 13 is interesting too. According to the pattern model V(a) = 3.01 Hh for Chord A and V(a) = 2.92 Hh for Chord B. This is a difference of 3% and possibly too small to affect a perceptual difference. Interestingly, in order *a* – *b* 7 participants preferred chord A and in order *b* – *a* 8 participants preferred the chord B. This finding is in support with the earlier result that overall an order effect had been observed.

4.3.4 Order Effects

As mentioned above, a significant order effect had been observed. It was particularly striking that trials in the order *a* – *b* received a more homogeneous rating in favor of the A chords, while the order *b* – *a* produced generally more noisy data. The question is why?

The author suggest to accept for now that the evidence of this experiment is clearly in favor of he pattern model indicating that especially in the context of uncommon chord the temporal model fails to make any predictions. Now, under the assumption, that the temporal model offers the "wrong" choice in form of the bass note, we could explain the order effect in terms that a confusing signal will effect the judgment of a participant in an experiment. The explanation is of the kind: If you present a participant of an experiment with the right choice first, (s)he pretty securely rejects the wrong choice when it is presented after the right choice. If you, however, give her/him the wrong choice first, (s)he end up being confused and goes more often for the wrong choice.

5. CONLUSION

As much as it had been shown that the pattern model as introduced by Hofmann-Engl (1990) had strong empirical validity, that it had been successfully implemented within contemporary composing (Hofmann-Engl, 1999) and that he can be a powerful tool in the analysis of contemporary classical music (Hofmann-Engl, 2004), the model had actually never been compared to any other model.

In this paper, this specific model was compared with Meddis's and Hewitt's temporal model (1991) and it was shown that the pattern model outmatched the temporal model predicting all roots to uncommon chords correctly.

Additionally, it became further clear that familiarity and context are powerful root predictors outweighing the roots as predicted by either model.

Finally, it has been demonstrated that order effects are strong in the context of virtual pitch experimentation.

It must be the goal of future research to find suitable methods to test common chords and to investigate the familiarity effect, context dependency and the order effect more systematically.

ACKNOWLEDGMENTS

A 'Thank you' to Ray Meddis for his collaboration during the year 2004. This project would not have been possible without the many discussions and the financial support from Essex University

REFERENCES

Hofmann-Engl, L. (1990). *Sonanz und Virtualität.* MA Thesis, TU Berlin (unpublished)

Hofmann-Engl, L. (1999). *Virtual Pitch and Pitch Salience in Contemporary Composing.* Proceedings of the VI Brazilian Symposium on Computer Music at PUC Rio de Janeiro

Hofmann-Engl, L. (2004). Virtual Pitch and its Application to Contemporary Harmonic analysis. Online Publication: www.chameleongroup.org.uk/research/link_virtual_analysis.html

Licklider, J. C. R. (1959). *Three auditory theories.* In: Psychology. A study of a Science (ed. Koch), Mac_Graw Hill, New York

Meddis, R. & Hewitt M. J (1991). *Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Pitch Identification.* J. Acoust. Soc. Am. 89, 2866 -2882

Moore, B. C. J. (1997). *An Introduction to The Psychology of Hearing*, Academic Press, 1997

Terhardt, E. (1976). *Ein psychoakustisches begründetes Konzept der musikalischen Konsonanz.* Acoustica, 36, 121-137

Terhardt, E. (1977). *The two-component theory of musical consonance.* In: Psychophysics and Physiology of Hearing (ed. Evans & Wilson)

Terhardt, E. (1979). *Calculating virtual pitch*, Hearing Research 1, 155-182

APPENDIX

Matlab Script for the temporal approach with minor modifications (Meddis, 2004)

```
function matchAMSpitch
disp('processing wav file')
wavFileName='filename.wav'
[signal sampleRate]=wavread(wavFileName);
fileLength=length(signal);
pars=[ ...
        ' FILENAME.dataFile_In.fileIn      ' wavFileName ...
    ];

[SACF info]=runDSAMsim('ACFfileIn.sim');
figure(1)
set(gcf,'position',[20 40 605 160])
plot(SACF)

expocount=1;
semitones=10.^[log10(1):   (log10(2)-log10(1))/12:   log10(2)];
F0s=55*semitones;
harmonics=[1:5]

for F0=F0s
    fprintf('processing F0= %6.1f  ', F0)
    dt=1/sampleRate;
    frequencies=F0* harmonics;
    toneDuration=0.1;
    dBSPL=0;
       initialSilenceDuration=0;    endSilenceDuration=0; rampDur=.005;
    signal=maketone(dt, frequencies, toneDuration, dBSPL, initialSilenceDuration, ...
        endSilenceDuration, rampDur);
    figure(2),  set(gcf,'position',[20 240 605 160])
    plot(signal)

    compareWavFileName='compare.wav';
    signal=normalize(signal);
    wavwrite( signal, 1/dt, compareWavFileName);

    pars=[ ...
            ' FILENAME.dataFile_In.fileIn  '   compareWavFileName ...
        ];
    [SACFcompare info]=runDSAMsim('ACFfileIn.sim', pars);
    plot(SACFcompare)

        x=sum(exp(-10(SACF-SACFcompare).^2));    expo(expocount)=x;
    disp(['expo= ' num2str(x)])
    figure(3), set(gcf,'position',[20 480 605 160])
    plot(F0s(1:expocount),expo)
    expocount=expocount+1;
    pause(1)
end

expo
[minexpo bestF0]= min(expo);
disp(['best F0 is ' num2str(F0s(bestF0))])
```