ICMPC SEOUL 1998

Proceedings of the Fifth International Conference on MUSIC PERCEPTION AND COGNITION

Seoul, Korea August 26-30, 1998

A Single Model Explaining the First and Second Pitch Shift Effects as Alternative Manifestations of a Single Phenomenon

PANTELIS N. VASSILAKIS
UCLA, Systematic Musicology, Music Perception & Acoustics Laboratory.
Box 951616, Los Angeles, CA 90095.
E-mail: pantelis@ucla.edu

The first pitch-shift effect

The *first pitch-shift effect* describes changes in the perceived pitch of a complex harmonic stimulus resulting from the uniform frequency shift Δf of all its components. De Boer (1956) conducted one of the first systematic studies examining this phenomenon, with stimuli consisting of 5 or 7 components spaced at f_0 = 200Hz and shifted by an equal amount of ΔfHz. The perceived pitch of each stimulus followed a saw-tooth-like function around the fixed frequency spacing (f_0), across the range of center frequencies (**Fig. 1**).

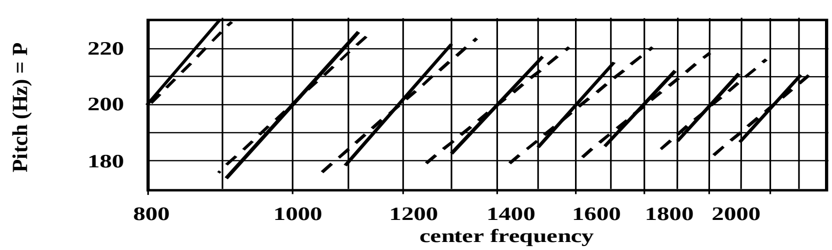


Fig. 1. (After de Boer, 1956). Dependence of the perceived pitch (**P,** abscissa) of a complex stimulus, with 5 equally spaced (f_0 = 200Hz) components, upon the frequency (ordinate) of its center component.

The continuous lines describe the observed relationship between pitch and center frequency while the broken lines describe the same relationship as predicted by the model in **Eq. 1** below. [**n**: component number corresponding to the center frequency of the nearest harmonic tone, and **d*f***: uniform frequency shift of components away from the nearest harmonic tone].

$$\mathbf{P} = \boldsymbol{f_0} + \frac{\mathbf{d}\boldsymbol{f}}{\mathbf{n}}, \quad (\Delta P = \mathrm{d}f\ /\ \mathrm{n}) \quad \underline{\textbf{Eq. 1}} \qquad \textbf{a)}\ \ if\ \ 0 \le |\Delta f| < f_0/2\ :\ \mathrm{d}f = \Delta f,\ \ \mathrm{n} = \mathrm{n}$$

$$\textbf{b)}\ \ -f_0 \le \Delta f < -f_0/2\ :\ \mathrm{d}f = \Delta f + f_0,\ \ \mathrm{n} = \mathrm{n} - 1 \qquad \textbf{c)}\ \ f_0/2 < \Delta f \le f_0\ :\ \mathrm{d}f = \Delta f - f_0,\ \ \mathrm{n} = \mathrm{n} + 1$$

The conditions: $0 \le |\boldsymbol{\Delta f}| < \boldsymbol{f_0/2}$ and $\boldsymbol{f_0/2} < |\boldsymbol{\Delta f}| \le \boldsymbol{f_0}$ describe the discontinuity in **Fig. 1**, manifested in a symmetrical jump of the perceived pitch around the value of $\boldsymbol{f_0}$, as the frequency shift crosses the value: $|\boldsymbol{\Delta f}| = \boldsymbol{f_0} / \mathbf{2}$ (area of ambiguous pitch).

Since de Boer (1956) various relevant studies have been conducted, confirming more or less his results. In order to account for the fact that the absolute value of the observed pitch shift is consistently larger than $|\boldsymbol{\Delta}\mathbf{P}|$, all studies introduce some sort of adjustment to the model described by **Eq. 1**, supporting it by various theoretical arguments. Smoorenburg (1970) presented the most convincing argument, demonstrating that the center frequency of a stimulus is reduced by the presence of combination tones[1]. This results in a lower value for **n**[2] and therefore a larger absolute value for the predicted pitch shift ($\Delta P = \mathrm{d}f/\mathrm{n}$). **Eq. 2** is a modification of **Eq. 1** based on this argument, predicting pitch-shift values that agree with observation.

$$\mathbf{P'} = \boldsymbol{f_0} + \frac{\mathbf{d}\boldsymbol{f}}{\beta\mathbf{n}} \quad (\Delta P' = \mathrm{d}f\ /\beta\,\mathrm{n}) \quad \underline{0 < \beta < 1} \quad \underline{\textbf{Eq. 2}} \qquad \text{(Same conditions as in Eq. 1)}$$

The new parameter, $\boldsymbol{\beta}$, represents the proportion of a spectrum that can be attributed to combination tones. In light of Smoorenburg's arguments, the actual value for β can be obtained empirically as the ratio:

$$\frac{\boldsymbol{\Delta P}\ \textit{(theoretical pitch shift)}}{\boldsymbol{\Delta P'}\ \textit{(experimental pitch shift)}} = \boldsymbol{\beta}$$

Schouten *et al*. (1962) used the data from a number of experiments and calculated an average value for this ratio to be: $\boldsymbol{\beta = 0.8}$[3]. Since the present study will use stimuli similar to those used in the above calculation, the value: $\beta = 0.8$ will be adopted whenever **Eq. 2** is used to predict the first pitch-shift effect.

The second pitch-shift effect

This term has been used to describe two distinct observations. The already discussed discrepancy between theoretical and experimental values of the pitch shift is one of these observations and, as we have seen, it does not constitute a new pitch-shift effect. The present study will use the term '*second pitch-shift effect*' to refer only to the following observation:
Changing the frequency spacing between the components of a harmonic stimulus, while keeping the center frequency fixed, results in opposite changes in pitch (**Fig. 2**).

[1] Smoorenburg referred to combination tones: $\boldsymbol{f_1} - \mathbf{k}\,(\boldsymbol{f_2} - \boldsymbol{f_1})$. His results indicated that combination tones contribute to the pitch of (in)harmonic tone complexes, *just as if those frequencies were present in the stimulus.*

[2] **n** = (center frequency) / (frequency spacing). As demonstrated by Smoorenburg, **n** may take values smaller than the number of the lowest component present in the original stimulus.

[3] They actually calculated the value: $b = (1/\beta) - 1$ and used it in a manner different than the present study.

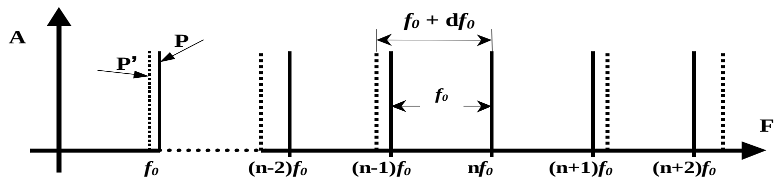


Fig. 2. Increasing the spacing of the components around a fixed center component (**n**), results in a drop in pitch **P'< P** although the difference frequency has increased from f_0 (continuous lines) to f_0+df_0 (broken lines).

All mathematical models that have been introduced until now to explain this observation include the parameter **df_0** (changes in the frequency spacing between components) in the calculations. They all predict that, for small values[4] of **df_0** the pitch will *always* drop when the frequency spacing between components is increased and rise when the frequency spacing is decreased. **i.e:**

(1) Schouten et al. (1962)

$$P = f_0 + \frac{(1+b)df}{n} - b\,df_0 \quad (b \approx 0.25)$$

(2) Schroeder (1966)

$$P = \frac{(nf_0 + df) - (f_0 + df_0)}{n-1}$$

In both of the above models, ***P*** and **df_0** are expected to move in opposite directions since the sign of **df_0** is being reversed.

Eq. 2, however, can predict the second pitch-shift effect without introducing the parameter **df_0**. Let us, for example consider the following pair of stimuli:

		5f_0	6f_0	7f_0	8f_0	9f_0	
stimulus-pair #1:	***Stimulus 1_a :***	950	1140	**1330**	1520	1710	f_0 = **190**Hz
	Stimulus 1_b :	930←	1130←	**1330**→	1530→	1730	f_0' = **200**Hz

Stimulus 1_a is harmonic with pitch: ***P(1_a)*** = f_0 = **190**Hz and *Stimulus 1_b* results from increasing the frequency spacing in *Stimulus 1_a* by 10Hz (f_0' = **200**Hz, ***P(1_b)*** = ?).

Let us now consider *Stimulus 1_b* as deriving *not* from *Stimulus 1_a* but from a harmonic tone (*Stimulus 1_{bb}*) with *frequency spacing*: f_0' = **200**Hz and *center frequency*: **1400**Hz, after shifting each component *down* by **Δ*f*** = **-70**Hz.

		5f_0'	6f_0'	7f_0'	8f_0'	9f_0'	
f_0' = **200**Hz	***Stimulus 1_{bb} :***	1000	1200	**1400**	1600	1800	***P(1_{bb})*** = **200**Hz
Δ*f* = **-70**Hz	***Stimulus 1_b :***	930	1130	**1330**	1530	1730	***P(1_b)*** = **187.5**Hz

In that case we can calculate the expected pitch of *Stimulus 1_b* using **Eq.(2)** with **n** = **7**, $f_0 = f_0'$ = **200**Hz, **d*f*** = **Δ*f*** = **-70**Hz and β = **0.8**:

4 No model explains why this should only happen for values of **df_0** that are small relatively to f_0 and no model predicts what will happen for larger values of **df_0**.

$$P(1_b) = f_0 + \frac{\mathbf{d}f}{\beta \mathbf{n}} = 200 + \frac{(-70)}{0.8 * 7} = 200 - 12.5 = \mathbf{187.5}Hz$$

As we can see, **Eq. 2** does predict a drop in pitch (*190Hz > 187.5Hz*) when increasing the frequency spacing between the components of *Stimulus* 1_a from 190Hz to 200Hz. This prediction has been confirmed repeatedly in all experiments that have examined the first pitch-shift effect (i.e. **Fig. 3**).

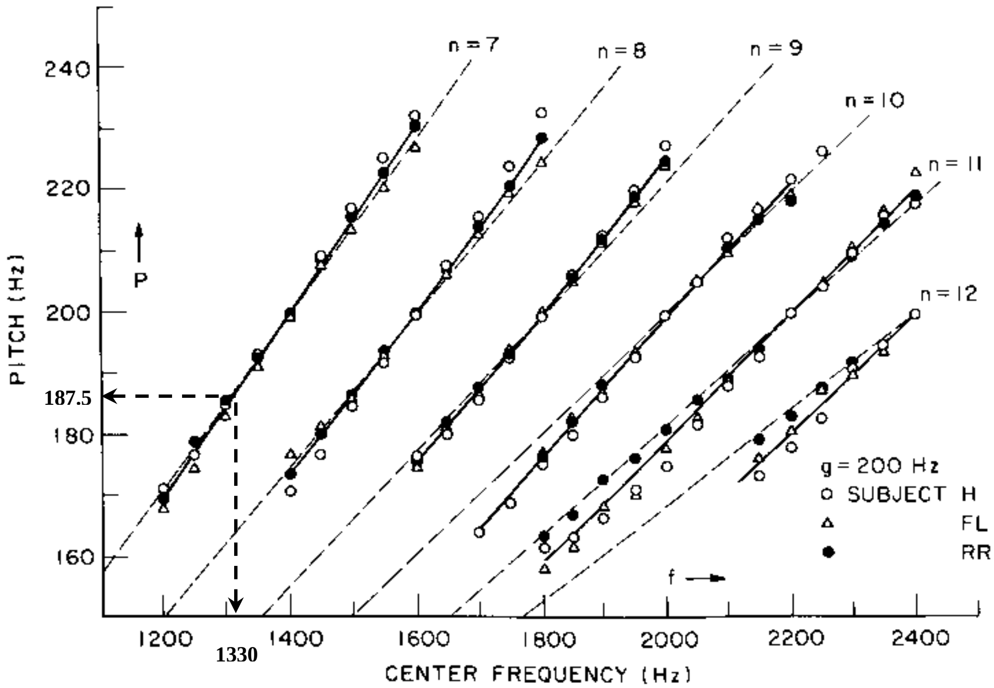


Fig. 3. (from Schouten, 1962). Pitch as a function of center frequency for 3-component stimuli. Solid lines represent the best fit of the experimental points. For **Center frequency** = **1330**Hz, **Pitch** = **187.5**Hz.

The reported drop in pitch when increasing the frequency spacing between components of a complex stimulus can therefore be predicted by the model of, as well as the data on the first pitch-shift effect.

New prediction

Describing the second pitch-shift effect in terms of **Eq. 2** means that the model's predictions are independent of the value $\mathbf{d}f_0$. This has an interesting implication. Suppose that, starting from *Stimulus* 1_b, we increase again the frequency spacing between its components from 200Hz to 210Hz:

<u>**stimulus-pair #2:**</u>

Stimulus $\mathbf{1_b}$ **:**	930 1130 **1330** 1530 1730	f_0 = **200**Hz
Stimulus1 $\mathbf{_c}$ **:**	910 ← 1120 ← **1330** → 1540 → 1750	f_0' = **210**Hz

As already shown, the predicted and observed pitch of *Stimulus* 1_b is: $\boldsymbol{P(1_b) = 187.5}Hz$.

To predict the pitch of *Stimulus* 1_c, let us again consider it as deriving *not* from *Stimulus* 1_b but from a harmonic stimulus (*Stimulus* 1_{cc}) with *frequency spacing*: $f_0' = \mathbf{210}$Hz and *center frequency*: **1260**Hz, after shifting each component *up* by $\Delta f = \mathbf{70}$Hz.

		$4f_0'$	$5f_0'$	$6f_0'$	$7f_0'$	$8f_0'$	
$f_0' = \mathbf{210}$Hz	***Stimulus*** 1_{cc} **:**	840	1050	**1260**	1470	1680	$P(1_{cc}) = \mathbf{210}$Hz
$\Delta f = \mathbf{70}$Hz	***Stimulus*** 1_c **:**	910	1120	**1330**	1540	1750	$P(1_c) = \mathbf{224.6}$Hz

We can now again calculate the expected pitch of *Stimulus* 1_c using **Eq. 2** with **n** = **6**, $f_0 = f_0' = \mathbf{210}$Hz, $\mathbf{d}f = \Delta f = \mathbf{70}$Hz and $\beta = \mathbf{0.8}$:

$$P(st1_c) = f_0 + \frac{\mathbf{d}f}{\beta\mathbf{n}} = 210 + \frac{70}{0.8 * 6} = 210 + 14.6 = \mathbf{224.6}Hz$$

In other words, **Eq. 2** predicts *a rise* rather than drop in pitch when increasing the frequency spacing between the components of *Stimulus* 1_b from 200Hz to 210Hz *(224.6Hz > 210Hz).*

This prediction is incompatible with the current definition of the second pitch-shift effect and contradicts all its existing models, which predict that frequency spacing and pitch will always change in opposite directions. At the same time none of the existing studies on the second pitch-shift effect contradict this prediction, since they have all been limited to starting stimuli that were harmonic (*i.e. stimulus-pair #1*). **Fig. 4**, below, displays the relationship between pitch and frequency spacing (second pitch-shift effect) as predicted by **Eq. 2**, for *center frequency = 1200Hz.* According to this model the direction of pitch change (rise/drop) for any two stimuli depends on their relative position within the graph rather than on prescribed changes in their frequency spacing (increasing/decreasing – large/small).

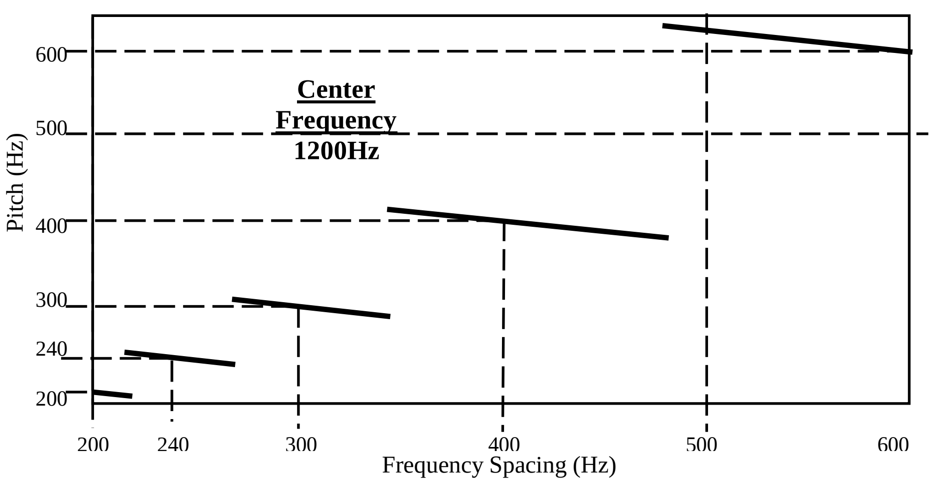
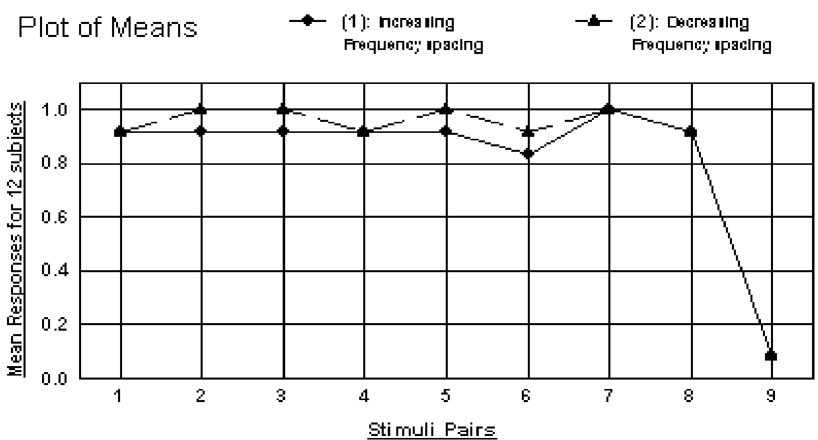


Fig.4. Second pitch-shift effect as predicted by Eq.(2), for center frequency: 1200Hz.

Experiment 1

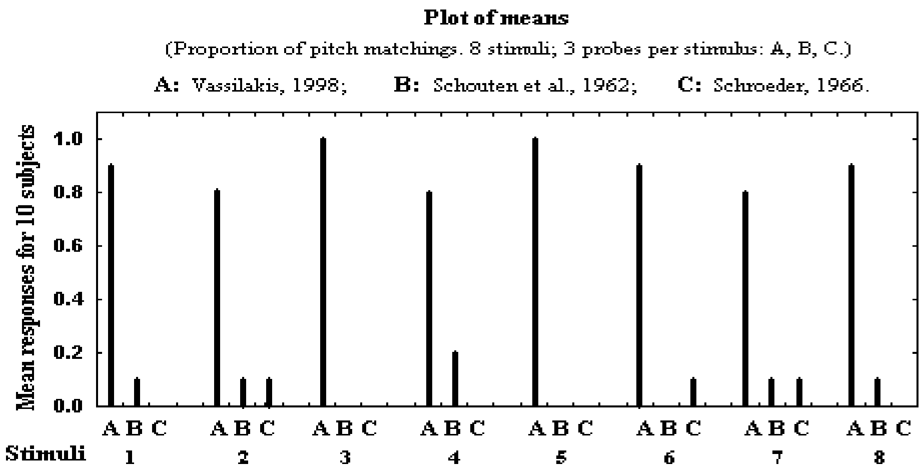
Nine pairs of complex tones were generated in mono, with 5 or 7 components each, frequency spacing between 150Hz and 220Hz, center frequency between 990Hz and 1540Hz, and changes in the frequency spacing between 10Hz and 20Hz. All tone-pairs were presented twice: **a → b**: *increasing frequency spacing*, **b→ a**: *decreasing frequency spacing*. Twelve subjects listened binaurally to the total of 18 tone-pairs in random order, and were asked to determine whether the pitch was going up or down when moving from the first to the second tone. Responses indicating that frequency spacing and pitch move in the same direction (new prediction) were coded as (1) and responses indicating the reverse (old prediction) were coded as (0). For the tone pairs included in the experiment, Eq.(2) predicted that frequency spacing and pitch will change in the *same direction* for *pairs 1-8* (expected response = 1) and in *opposite directions* for *pair 9* (expected response = 0). The mean responses across subjects are plotted below:



Two-way multinomial ANOVA indicated a zero 2-way interaction and no main effect for Factor 1 (direction of changes in frequency spacing). Post-hoc analysis (*Scheffé, a = 0.05*) on the main effect of Factor 2 (stimuli-pairs) supported the model's predictions.

Experiment 2

Eight complex tones were generated in mono and their components were both shifted and stretched (or contracted) away from harmonic relationships so that *Eq. 2*, *Schouten's model*, and *Schroeder's model* predicted different pitches. Ten subjects were presented randomly with all eight stimuli and were asked to match them in pitch with one of three probes generated based on the three models. The results, plotted below, indicated a significantly higher degree of reliability for the model introduced by the present study (Eq. 2).



Conclusions

1. The relationship between pitch and frequency spacing among the components of a complex stimulus (no requirement for fixed center frequency, small changes in frequency spacing, or opposite motion between pitch & frequency spacing) can be predicted by the same model that predicts the relationship between pitch and center frequency of a complex stimulus with equally spaced components. First and second pitch-shift effects can therefore be considered alternative manifestations of the same phenomenon, the pitch-shift effect.

2. Eq.(2) is a powerful model of the pitch-shift effect. It allows us to predict the pitch of any complex stimulus with equally spaced components regardless of frequency spacing, center frequency, or changes in frequency spacing and center frequency combined.

References

de Boer, E. (1956). *On the 'Residue' in Hearing.* Doctoral dissertation. The Netherlands: University of Amsterdam.

Schroeder, M. R. (1966). Residue Pith: A Remaining Paradox and a Possible Explanation. *JASA*, 40(1): 79-81.

Schouten, J.F., Ritsma, R. J. & Cardozo, B. L. (1962). Pitch of the Residue. *JASA*, 34(8/2): 1418-1424.

Smoorenburg, G.F. (1970**)**. Pitch Perception of two-frequency Stimuli. *JASA*, 48(4/2): 924-942.