The pitch of short-duration vibrato tones: experimental data and numerical model

C d'Alessandro & M Castellengo

Abstract

Vibrato is one of the most common ornaments in occidental classical music, particularly in singing. For actual vibrato tones, performed in a real musical situation, the tone durations are limited according to the musical score. The aim of the present work is to provide experimental data on pitch perception for vibrato tones of short duration. The pitch perceived of short vocal vibrato tones was measured using a method of adjustment. The stimuli were synthetic vocal tones, produced by a formant synthesizer. The main parameter under study was the tone duration, as a function of the fractional number of vibrato cycles. This parameter was examined in relation to: 1) vibrato extent (0, 50, 100 and 200 cents); 2) vibrato rate (4, 6 and 8 Hz); 3) nominal frequency (220, 440, 880 and 1500 Hz); 4) initial phase of the vibrato waveform (0, $\pi/2$, π, $3\pi/2$), assuming a sinusoidal vibrato waveform. Durations ranging from 1/2 cycle to 2 cycles were studied. A group of 12 musically educated subjects was used for the main experiment (nominal frequency 440 Hz, vibrato frequency 6 Hz, vibrato extent 100 cents). A smaller group of 4 selected subjects was used for the parametric experiments. Our results demonstrate that for short tones the pitch does not correspond to the mean frequency between the peaks of F0, as seems to be the case for long tones: rather, it corresponds to a weighted time-average of the F0 pattern. A separate perception took place for the high and low parts of the vibrato cycles, for large vibrato extents or slow vibrato rates. This phenomenon was related consistently with the glissando threshold. A simple numerical model of weighted time-averaging with threshold is proposed. It demonstrated a good agreement with our experimental data. Finally, the experimental results obtained and the model proposed provide some insight into questions related to vibrato in actual singers' performances: synchronization of vibrato cycles with note transitions, vibrato patterns in large pitch changes (portendo etc.), vibrato patterns in short staccato tones.

1.Introduction

In actual singing, the durations of vibrato tones are limited. This induces an interaction between the intrinsic vibrato parameters and the tone patterns imposed by the musical score. This interaction is responsible for varied and complicated fundamental frequency (F0) patterns, that are noticeable on analyses of actual singers performances. Understanding pitch perception for short-duration vibrato tones appears as a challenging problem for music perception studies.

In an extensive study by Shonle and Horan (1980), it was shown that the pitch which is perceived for long vibrato tones is the mean between the two extreme frequencies, with the geometric mean being a somewhat better fit than the arithmetic. In their experiments the number of vibrato cycles was indefinite, and no attention was paid to the initial and final phases of the modulation patterns. These results can hardly be extended to short tones. In this later case two additional parameters (tone duration and modulation phase) must be taken into account, in addition to the three intrinsic vibrato parameters studied for long tones (nominal frequency, extent and vibrato frequency).

In a previous work, we presented an experimental study on pitch perception for short duration vibrato tones. Due to space limitation, the reader is referred to (d'Alessandro & Castellengo, 94) for details on the experiments. In this paper, we review these experimental data, and we question the possible numerical models of pitch perception for short-duration vibrato tones (geometric or arithmetic means, time-average, weighted time-average). Finally, musical examples are discussed in the light of our model.

2. Experimental data and results

2.1 Experimental data

The pitch for short-duration vibrato tones was measured using a method of adjustment, and synthetic vocal stimuli. The experimental conditions are described in (d'Alessandro & Castellengo, 94). Figure 1 shows the F0 patterns used in the experiments: two different final phases

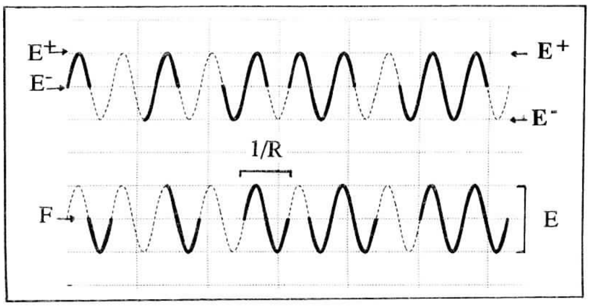


Figure 1: F0 patterns used in the experiments. E: vibrato extent; F: vibrato nominal frequency; R vibrato rate; E^+: F0 maximum; E^-: F0 minimum.

(0 and π) and seven durations (0.5, 0.75, 1, 1.25, 1.5, 1.75, 2 cycles) were studied. In the main experiment, the vibrato extent (E) was 100 Cents, the vibrato rate (R) was 6 Hz, and the nominal frequency (F) was 440 Hz. For parametric experiments, only five durations were studied (0.5, 0.75, 1, 1.5, 2 cycles). There were three parametric experiments: exp. P1: E was varying in four steps (0,50,100,200 Cents) with R=6 Hz and F=440 Hz; exp. P2: F was varying in four steps (220, 440, 880 and 1500 Hz) with R=6 Hz and E=100 Cents; exp. P3: R was varying in three steps (4,6,8 Hz) with E=100 Hz and F=440 Hz. Sound Example 1 consists of seven F0 patterns, with the same initial phase 0, the same vibrato rate 6 Hz, the same nominal frequency 440 Hz, the same vibrato extent 100 cents, and with seven durations 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2 cycles. This example demonstrates that pitch depends on the shape of F0 patterns.

Figure 2 shows the results obtained (2A: main experiment; 2B,C,D: parametric experiments). X axis represents the pattern duration. Y axis represents the difference in Cents between the pitch judgement (average of all subjects responses) and the nominal frequency. It must be pointed out that for the 0.5 cycle conditions, the F0 pattern is always either above or below the nominal frequency (see Figure 1, first pattern on each line): this point will be discussed in the next section. Patterns sharing the same final phase are connected by a similar line. The two final phases, 0 and π that are present in our stimuli are therefore represented by two different lines.

2.2 F0 integration and influence of the final phase

It is noticeable that the lines are converging towards the center frequency. Figure 2 demonstrates that the pitch judgements were influenced most by the shape of F0 patterns at the end of the tones. The experiment P2 (Figure 2D) indicates that the convergence toward the center frequency is almost independent of this center frequency: the lines for 220, 440, 880, and 1500 Hz are close together, when expressed in Cents relative to the nominal frequency.

In experiment P1 (Figure 2C), the lines clearly converge towards the center frequency for vibrato extent 0, 50 and 100 Cents. For the 200 cents extent condition, all the tones were perceived as either significantly higher or lower than the nominal frequency, almost independantly of their duration. This indicates that a separate perception of the high and low parts (extrema) of the F0 patterns appeared: the stimuli were perceived as two alternating tones rather than vibrato tones. Borrowing the terminology introduced in Nabělek and al. (1970), in their study of pitch for tone bursts of changing frequency, we shall adopt the term "separation" when only the final arch of the vibrato waveform contributes to the pitch judgement, and the term "fusion" when the overall F0 pattern contributes to the pitch judgement. In a musical context, fusion could correspond to true vibrato, and separation might rather be related to trills or other melodic ornaments. Therefore, if the vibrato extent is large, a separate perception of the upper and lower arches of the vibrato waveform occurs (the final arch contributes alone to pitch perception, the overall F0 pattern has very little influence, if any, on the pitch perceived). In the case of fusion, although the final part of the F0 pattern seems still the most important for pitch perception, the overall F0 pattern contributes to the pitch judgement.

For experiment P3 (Figure 2B), the same F0 patterns for different vibrato frequencies have different durations, because they have the same number of cycles. For a slow vibrato rate (4 Hz) all the F0 patterns are perceived either high or low. Therefore, separation occurs for a slow vibrato rate (as was the case for a large extent), and in the case of fusion, the lines are converging faster for the higher vibrato rate. This observation is compatible with the observation that the overall F0 pattern contributes to pitch perception in case of fusion.

2.3 Fusion/separation and the glissando threshold

As the fundamental frequency changes continuously in vibrato tones, the fusion/separation problem is related to the absolute threshold of pitch change, or glissando threshold. A unified view of this problem was

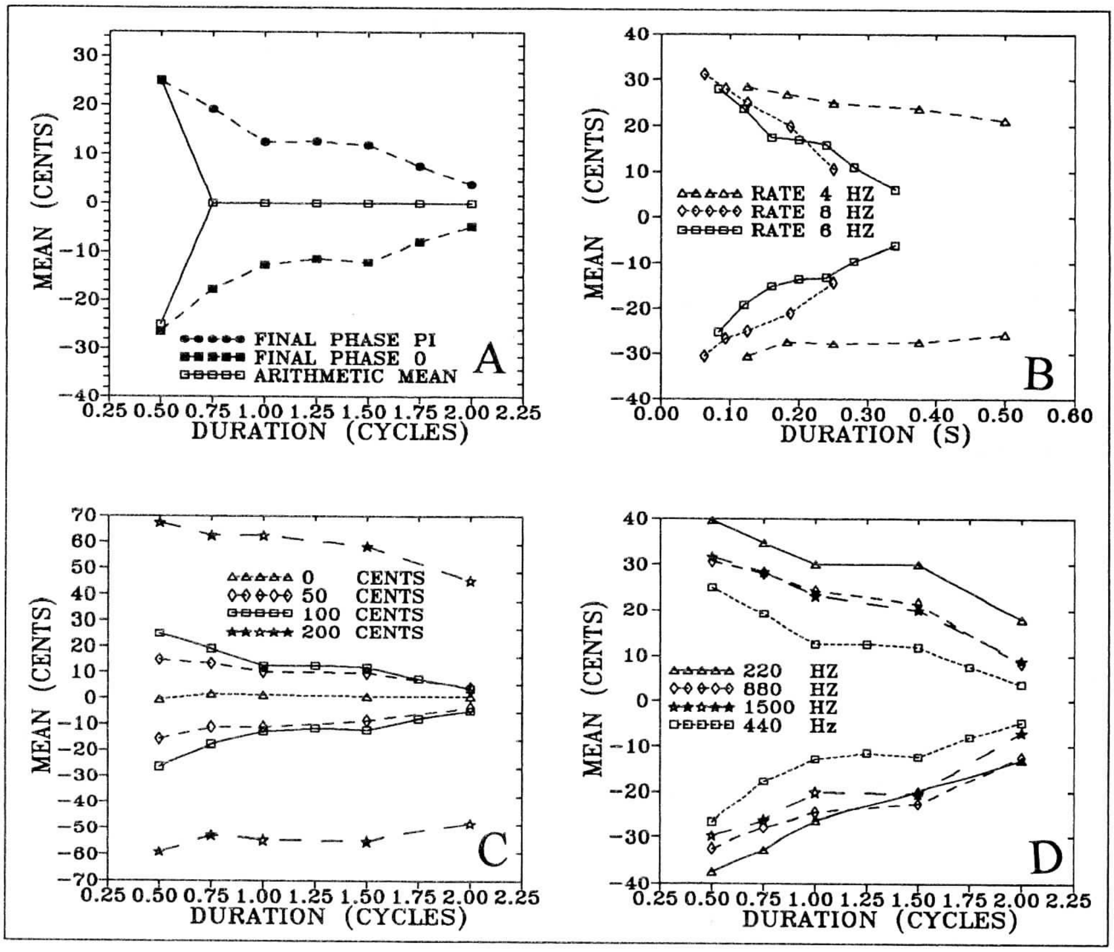


Figure 2: Experimental data. A: main experiment and mean frequency model; B: P3; C: P1; D: P2.

presented by 't Hart and al. (1990). Interpreting 't Hart data in the case of vibrato tones, one can expect a separated perception if E * R^3 > 0.02. The fusion/separation situations observed in the experiments P1 and P3 are in good agreement with the glissando threshold. The glissando rates in case of separation are close to the glissando threshold, although they are clearly below the threshold in case of fusion.

3. Numerical models

3.1 Mean Pitch Model

It has been generally accepted that the pitch which is perceived for long tones with vibrato is the mean pitch, with the geometric mean being a somewhat better fit than the arithmetic. Following a question raised by Sundberg (1993), we shall discuss this mean pitch hypothesis for short tones. The mean pitch is computed as the arithmetic or geometric mean between the two extreme values E^+ and E^- of the F0 excursion (see Figure 4 top).

The mean pitch equals the nominal frequency for all the patterns used in our experiments, excepted the 0.5 cycles, phase 0 or π patterns, where the mean pitch equals the nominal frequency± half the vibrato extent expressed in Hz (the mean pitch for the main experiment is plotted in Figure 2A.). This is the reason why we chose to plot the experimental results as the difference in Cents compared to F, instead of a difference in Cents compared to the mean pitch. The 0.5 cycles phase 0 or π conditions are the only conditions where there is coincidence between the mean pitch and the perceived pitch: it must be pointed out that in these particular conditions other models (e.g. time average or weighted time average) give almost the same result. Moreover the mean pitch hypothesis is unable to explain all the other data, because all the other patterns have the same mean pitch (which equals the nominal frequency). Finally the mean pitch hypothesis does not take into account the fusion/separation phenomenon which emerges from the experimental data. Figure 4 top

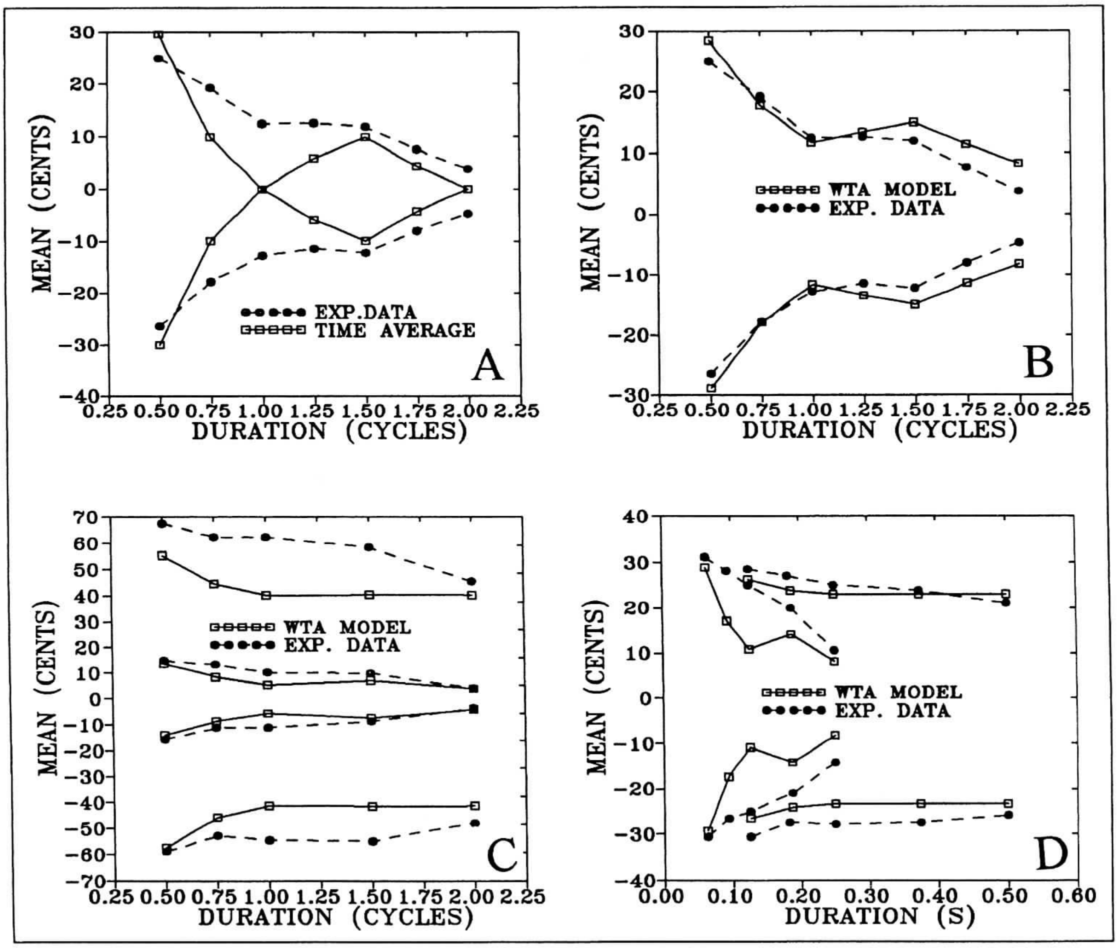


Figure 3: Models. A: time-average, main experiment. B: WTA, main experiment. C: WTA, P1. D: WTA P3.

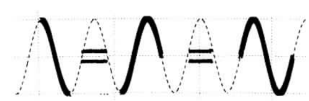
demonstrates the inadequacy of the mean pitch hypothesis, showing F0 patterns that are perceived with different pitches, but that have the same arithmetic or geometric means.

3.2 Time-Average Model

For long tones, Shonle & Horan (1980) also proposed a subsidiary experiment, using asymetrical vibrato waveforms. The results obtained indicated an averaging of all frequencies present, rather than a mean between the extreme frequencies. Let p(t) denote the pitch perceived at time t, and f the time-varying F0 function begining at time t. The equation for the time-average model is plotted in Figure 4 middle. But a simple time-

Means

$$p(t) = \begin{cases} (E^+ + E^-)/2 & \text{Arithmetic Mean} \\ \sqrt{E^+ \times E^-} & \text{Geometric Mean} \end{cases} \quad (1)$$



Time Average

$$p(t) = \frac{1}{t}\int_0^t f(\tau)d\tau \quad (2)$$

Weighted Time Average

$$p(t) = \frac{\int_0^t (exp(-\alpha(t-\tau)) + \beta) f(\tau) d\tau}{\int_0^t (exp(-\alpha(t-\tau)) + \beta) d\tau} \quad \text{with} \quad (3)$$

$$\begin{cases} \alpha = -22 \;\; \beta = 0.2 & \text{if } [E^+ + E^-] \times R^3 > 0.02 \\ \alpha = -22 \;\; \beta = 0 & \text{if } [E^+ + E^-] \times R^3 \leq 0.02 \end{cases}$$

Figure 4: Equation of the models. Equal responses for different patterns which are perceived with different pitches are indicated.

average model is no more sufficient to explain our data, because for an integer number of vibrato cycles the time-average equals the nominal frequency, although the pitch judgements we obtained in these cases are not the nominal frequency. Another argument against the time-average model is the monotonic convergence: towards the nominal frequency observed in case of fusion. Figure 3A shows the time-average model output for the main experiment, and demonstrates its inadequacy. Figure 4 middle shows some F0 patterns that are perceived with different pitches, but that have the same time-average.

3.3 Weighted Time-Average Model

It appeared that the final part of the tone had a larger weight on the pitch judgement than the initial one. A quantitative model for such a process is a time-average of the F0 pattern viewed through a data window. A simple model for the data window is a raised exponential memory function, so that events in the past contribute exponentially less to the average. Let α, β be two constants, Figure 4 bottom gives the equation of the weighted time average (WTA) model. The constant β accounts for the amount of long term time averaging, and the constant α accounts for the weighting of the past (speed of decay of the exponential function). In the case of separation, the excitation patterns due to successive vibrato extrema become more separated in time or in frequency: it is reasonable to assume that if the glissando rate for a given condition is larger than the glissando threshold (separation), the amount of constant time-averaging represented by β is reduced.(β= 0).

Only two free parameters α and β have to be estimated. Optimal parameters have been estimated by minimizing the Root Mean Square distance between the response of the model and the experimental data. The minimum distance is reached for α=-22 and β=0.20.(20 % of long term time averaging). This numerical model seems to be well-suited to both fused and separated vibrato judgements, as it appears on Figure 3B, 3C, 3D.

5. Musical examples

Our results demonstrate that for short tones the pitch does not correspond to the mean frequency between the peaks of F0, as it seems to be the case for long tones. We found that for short-duration vibrato tones, the pitch perceived corresponds to a weighted time-average of the F0 pattern. We shall examine now two musical examples in the light of this model. The first example, Figure 5 Top (sound example 2) shows the spectrogram of a

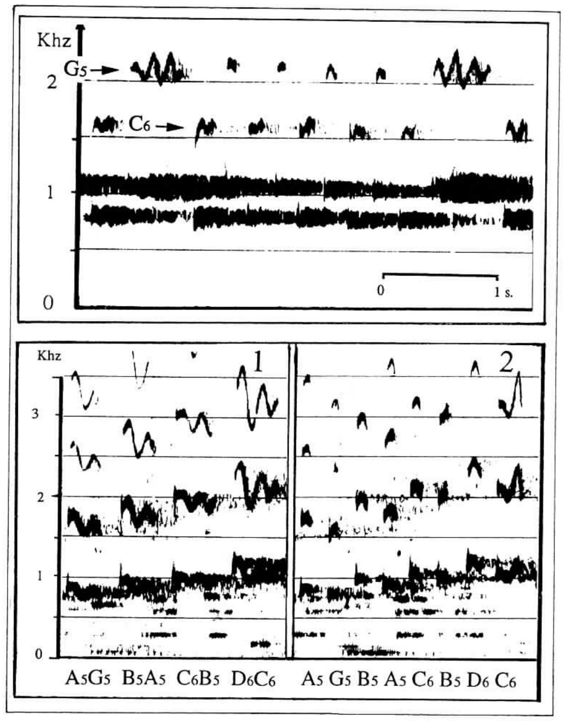


Figure 5: Musical examples. Top: "la fille du régiment" (Donizetti, Joan Sutherland). Bottom: "O légère hirondelle" (Gounod, Mado Robin).

sequence extracted from "La fille du régiment" (Donizetti) sung by Joan Sutherland, LP London OSA 13107. Various F0 patterns are present in the two alternating short tones of this virtuoso passage. As predicted by our model, one can notice that the different vibrato patterns for a same pitch do have different nominal frequencies, according to their durations, rate and extent.

The second example, Figure 5 Bottom (sound example 3) shows the spectrogram of two sequences extracted from the Air "O legère Hirondelle" (Mireille, Gounod), sung by Mado Robin, LP Decca LXT 2898. The two sequences consisted of variations on the same notes (legato by two, then staccato). Let us consider the last vibrato pattern of each example. Our model is able to explain the fact that in the first case (1) two notes are perceived, and in the second case (2) only one note is perceived. The glissando threshold is about 34 Semi-Tone/s for both. In (1) the glissando rate is about 53 ST/s at the begining (separation between D6 and C6). For (2) the rising glissando rate is about 38 ST/s, of the same ordre of magnitude than the threshold (there is fusion, only one note C6 is perceived).

References

d'Alessandro C. & Castellengo M. (1994), "The pitch of short-duration vibrato tones.", J. Acoust. Soc. Am., 95, 1617-1630.

Nabelek I V, Nabelek A. K & Hirsh I.J, (1970), "Pitch of tone bursts of changing frequency", J. Acoust. Soc. Am. 48, 536-553.

Shonle J.I, and Horan K.E,.(1980), "The pitch of vibrato tones", J. Acoust. Soc. Am. 67, 246-252.

Sundberg, J, (1993), Personal communication at SMAC93.

'tHart J, Collier R, and Cohen A, (1990), "A perceptual study of intonation", (Cambridge Univ. Press, UK), pp. 29-33.