



**HAL**  
open science

## The pitch of short-duration vibrato tones

Christophe d'Alessandro, Michèle Castellengo

► **To cite this version:**

Christophe d'Alessandro, Michèle Castellengo. The pitch of short-duration vibrato tones. Journal of the Acoustical Society of America, 1994, 95 (3), pp.1617-1630. hal-02008510

**HAL Id: hal-02008510**

**<https://hal.science/hal-02008510>**

Submitted on 5 Feb 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## The pitch of short-duration vibrato tones

Christophe d'Alessandro, and Michèle Castellengo

Citation: *The Journal of the Acoustical Society of America* **95**, 1617 (1994); doi: 10.1121/1.408548

View online: <https://doi.org/10.1121/1.408548>

View Table of Contents: <https://asa.scitation.org/toc/jas/95/3>

Published by the *Acoustical Society of America*

---

### ARTICLES YOU MAY BE INTERESTED IN

#### [The pitch of vibrato tones](#)

*The Journal of the Acoustical Society of America* **67**, 246 (1980); <https://doi.org/10.1121/1.383733>

#### [The pitch of short-duration fundamental frequency glissandos](#)

*The Journal of the Acoustical Society of America* **104**, 2339 (1998); <https://doi.org/10.1121/1.423745>

#### [Frequency difference limens for short-duration tones](#)

*The Journal of the Acoustical Society of America* **54**, 610 (1973); <https://doi.org/10.1121/1.1913640>

#### [Vibrato extent and intonation in professional Western lyric singing](#)

*The Journal of the Acoustical Society of America* **102**, 616 (1997); <https://doi.org/10.1121/1.419735>

---

# The pitch of short-duration vibrato tones

Christophe d'Alessandro  
*LIMSI-CNRS, BP 133, F-91403 Orsay, France*

Michèle Castellengo  
*Laboratoire d'Acoustique Musicale, CNRS, Université Paris VI, Ministère de la Culture, Tour 65-66,  
Université Paris VI, 4 Place Jussieu, F-75005 Paris, France*

(Received 5 February 1993; accepted for publication 30 November 1993)

The pitch perceived for 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) the vibrato extent (0, 50, 100, and 200 cents); (2) the vibrato rate (4, 6, and 8 Hz); (3) the tone nominal frequency (220, 440, 880, and 1500 Hz). Durations ranging from  $\frac{1}{2}$  cycle to 2 cycles were studied. Our results showed that for short tones, the pitch perceived does correspond to a weighted time average of the  $F_0$  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 was proposed. It demonstrated a good agreement with our experimental data. Finally, the experimental results obtained, and the model proposed, were able to explain the musical perception of complex  $F_0$  patterns, like those present in actual singing.

PACS numbers: 43.75.Bc, 43.66.Hg, 43.75.Rs

## INTRODUCTION

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  $F_0$  contours ( $F_0$  denotes herein the fundamental frequency) of actual singers' performances are varied and complicated patterns. On the one hand some variation of  $F_0$  is imposed by the vibrato pattern (depending on vibrato frequency, extent and phase), and on the other hand, some variation of  $F_0$  is due to the tones and tone transition patterns (depending on tone frequency, duration and mode of articulation) as they are noted in the musical score. As a matter of fact, observing natural  $F_0$  contours Carl Seashore (1938, p. 50) noted as early as 1938 "... the astounding difference between the actual vibrato as it exists in the physical tone, and the vibrato as it is heard in the musical situation." This competition between vibrato patterns and tone patterns is a critical point especially for short tones (tones with a small number of vibrato cycles).

Let us examine an example, to give an idea of the constraints imposed on the  $F_0$  pattern in a real musical situation, by both the vibrato and the score. Figure 1 shows the spectrogram of a well-known short musical sequence (Mozart's Magic Flute, Aria of the Queen of the Night "Hell's vengeance seethes in my heart," sung by Beverly Hoch, CD EMI 7 54287 1/2/4/). Vibrato is present even in the very short tones of this virtuoso passage. This is an example of very good musical performance, and the accuracy in intonation is excellent. Nevertheless, it is not clear why the varied  $F_0$  patterns that are actually present in this

performance are perceived as the same musical note (the eight repeated eighth notes in the boxed score).

The aim of the present work is to provide experimental data, and a model derived from these experimental data, on pitch perception for short duration vibrato tones, like those in Fig. 1.

In an extensive study by Shonle and Horan (1980), it was shown that the pitch which is perceived for long vibrato tones is close to the geometric mean between the two extreme frequencies. In their experiments, long and indefinite vibrato tones were used: A carrier tone was frequency modulated, without paying any attention to the beginning and end of the tone. They also presented a review and a discussion of the literature on vibrato, to which the reader is referred for more information on previous work on vibrato perception for long tones. Shonle and Horan concluded that "... the pitch of vibrato tones for the frequency range 220 to 1500 Hz, with modulation widths up to a total of 200 Cents, is the mean pitch, with the geometric mean being a somewhat better fit than the arithmetic." This hypothesis will be called the mean between extreme (MBE) hypothesis. Let  $p$  denote the pitch perceived,  $f(t)$  the periodic time-varying  $F_0$  function, the MBE pitch is computed as  $[\text{Max}(f) + \text{Min}(f)]/2$  (arithmetic MBE) or  $\sqrt{[\text{Max}(f) \times \text{Min}(f)]}$  (geometric MBE). In this later study, vibrato perception was studied as a function of the modulation waveform, of the vibrato extent, of the vibrato rate and of the signal frequency. The tone duration was not a parameter under study.

In a study by Sundberg (1978) the pitch perceived for long synthetic vocal tones (1.9 s) was measured using a method of adjustment. Four fundamental frequency values were used (70, 115, 185, and 300 Hz), a modulation width

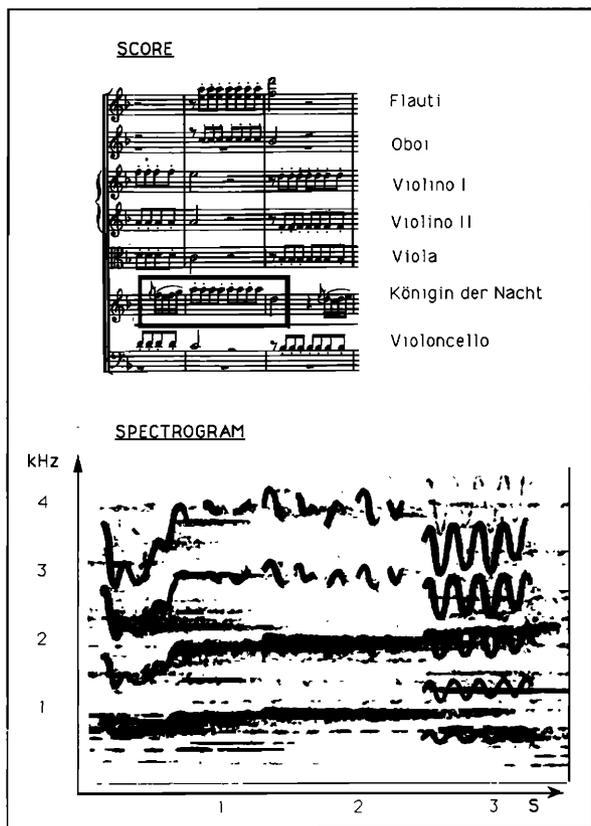


FIG. 1. Score and spectrogram of a short sequence extracted from Mozart's Magic Flute: Aria of the Queen of the Night "Hell's vengeance seethes in my heart," sung by Beverly Hoch, CD EMI 7 54287 1/2/4/.

of 70 cents, and a vibrato rate of 6.5 Hz. The author concludes that "The pitch perceived from a signal having a sinusoidal vibrato corresponds closely to a frequency which is the linear average of the time varying fundamental frequency." This hypothesis will be called the linear average (LA) hypothesis. Let  $p$  denote the pitch perceived,  $f$  the periodic time-varying  $F_0$  function with vibrato period  $T$ . For a long tone (without taking into account the tone duration), the LA pitch is computed over one complete vibrato period:

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

This definition can naturally be extended to short duration tones, computing the average over the tone duration  $t$ :

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

All the studies that we have been able to locate on pitch perception for vibrato tones concerned only long tones, with a large (and unspecified) number of vibrato cycles.

It appeared questionable whether the conclusions proposed for long tones can apply to short-duration vibrato tones, like those in Fig. 1 for instance. Both MBE and LA hypotheses give the same response for modulation patterns like No. 9 and 11, or No. 6 and 7, in Fig. 3, although they

look very different, and one could suspect that they could be perceived with different pitches. New experiments were therefore needed for short tones.

In a previous paper by Castellengo *et al.* (1989), some results indicating that the pitch perceived for short-duration vibrato tones can be far from the mean between the extreme frequencies were presented. Some experiments were reported in d'Alessandro and Castellengo (1991, 1992), and they are completed in the present article.

Experiments are described in which the pitch perceived for short vibrato tones was measured, taking the same terminology and the same parameter values as in Shonle and Horan (1980) for center frequency (220, 440, 880, 1500 Hz), vibrato extent (0, 50, 100, 200 cents), vibrato rate (4, 6, 8 Hz), and the same experimental paradigm (method of adjustment). In addition to these parameters, the new parameters under study are the  $F_0$  pattern durations and the  $F_0$  pattern phases, assuming a sinusoidal vibrato waveform. We shall report the results of these experiments together with a simple numerical model which is able to represent our data quite well.

## I. METHOD

### A. Experimental procedure

The aim of this study was to measure the pitch perceived for short-duration vocal vibrato tones. The method of adjustment seemed appropriate for our purpose because it has a number of advantages: (1) each response gives an estimation of the pitch perceived; (2) it helps concentration, as the subject is actively engaged in the adjustment process; (3) the data are less variable than forced-choice data in frequency discrimination tasks [see for instance the study by Wier *et al.* (1976)].

The stimuli were synthetic vowels /a/, digitally computed by a parallel formant synthesizer. Vocal synthetic stimuli were chosen because they are closer to natural stimuli than pure tones or square waves. It has been shown otherwise in Shonle and Horan (1980) that the pitch perceived does not show any strong dependence on complexity of the signal waveform, and therefore we preferred here vocal-like waveforms. The vibrato waveform was sinusoidal (with value 0 for a 0 phase).

The stimuli consisted of two tones of the same duration, with a 300-ms silent interval between each member of a stimulus pair. This silent interval duration seemed not critical. Time lines for this psychophysical procedure are shown in Fig. 2. The first tone ( $V$ ) for each stimulus pair was a vibrato tone, the same for all the stimulus pairs of a given condition, and the second tone ( $C$ ) had a constant frequency, the value of which varied among the different stimulus pairs.<sup>1</sup>

Subjects were instructed to adjust the frequency of the second tone ( $C$ ) to be equal to that of the first tone ( $V$ ). The adjustment procedure was not continuous, as the frequency of the ( $C$ ) tones varied in discrete steps. Typically, subjects could choose a response from 20–30 discrete frequencies, depending on the parameters involved. The frequency difference between the ( $C$ ) tones of two contiguous

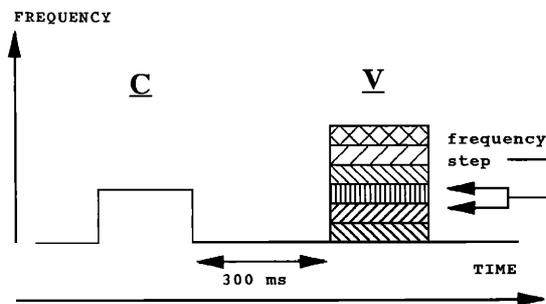


FIG. 2. Time line for the discrete-trial adjustment procedure: (V) vibrato tone, (C) varying constant- $F_0$  tone.

points on this adjustment scale was always less than the frequency difference limen for this center frequency. The adjustment procedure was therefore “psychologically continuous,” even if it was not “physically continuous.”<sup>2</sup> The test was computerized, and the subjects could adjust the frequency of the (C) tone using a mouse device to move a sliding bar on the computer screen. This procedure introduced a visual reference, which is a drawback of such an adjustment paradigm. Nevertheless, the results indicated no particular preference for the center of the sliding bar, and we can therefore hypothesize that this visual reference had very little influence on the subjects’ judgements. The signals were presented diotically at a level of 80 dB SPL, directly from the computer digital-to-analog 16-bit converter, through Beyer DT48 headphones.

For each pattern, the subjects could adjust  $F_0$  for the tone (C) in the limits of the peaks of frequency variation for the tone (V). The subjects were asked always to give an answer, even if the pitch seemed ambiguous. The same question was asked for all experiments: Adjust the pitch of the second tone of a stimuli pair in order to match the pitch of the first one. For each  $F_0$  pattern, the subjects had no time limits to give an answer. Subjects typically listened to 10 stimulus pairs while manipulating the frequency of tone (C) before reporting a match. The session duration was recorded, and the average duration of a test session for one experimental condition (about 30 matches) was around 15 min. This experimental procedure is somewhat

TABLE I. Combinations of center frequencies, vibrato extents, vibrato rates, duration for the preliminary experiments. Also shown are the average settings and standard deviations in cents, compared to the results of Shonle and Horan for a continuous vibrato tone. Duration = 10 cycles, 4 subjects, 20 matches.

Center frequency (Hz)	Vibrato extent (cents)	Vibrato frequency (Hz)	Results (cents)	Results S&H (cents)
220	100	6	$7.8 \pm 10.7$	$-0.3 \pm 8.0$
440	0	...	$1.1 \pm 2.8$	$-1.9 \pm 4.4$
440	50	6	$0.1 \pm 4.2$	$-1.4 \pm 5.8$
440	100	4	$3.9 \pm 14.9$	$2.5 \pm 9.7$
440	100	6	$-1.7 \pm 7.9$	$-2.9 \pm 8.0$
440	100	8	$-6.6 \pm 9.5$	$1.1 \pm 11.9$
440	200	6	$0.2 \pm 7.7$	$-6.4 \pm 14.2$
880	100	6	$3.9 \pm 9.5$	$-0.5 \pm 11.7$
1500	100	6	$1.9 \pm 5.9$	$0.6 \pm 3.5$

similar to the procedure used for long tones in the work by Sundberg (1978), which came to our attention only after completion of the work reported in this paper.

### B. Test conditions

The parameters under study were: The center frequency, the vibrato extent, the vibrato rate, the pattern duration, and the pattern initial phase (for a sinusoidal vibrato function). A preliminary experiment was designed, in order to check our experimental procedure, using 10 cycle tones (considered as long tones).

Table I contains the various combinations of parameter used in the preliminary experiment, along with the results, compared to those obtained in Shonle and Horan (1980). The total number of matches was 20 for each condition in this preliminary experiment.

The various combinations of parameters used in the other experiments, along with the number of subjects and the total number of matches for each condition are given in Table II. The duration represents a number of vibrato cycle rather than an absolute duration.<sup>3</sup>

We shall refer to condition 5 in Table II as the main experiment, and to the other condition as the parametric experiments. This condition is called “main experiment,”

TABLE II. Combinations of center frequencies, vibrato extents, vibrato rates, duration, initial phases, number of subjects, and total number of matches for the main (condition 5) and the parametric experiments.

	Center frequency (Hz)	Vibrato extent (cents)	Vibrato frequency (Hz)	Duration Number of cycles	Initial phase (radians)	Number of subjects	Number of matches
1	220	100	6	0.5, 0.75, 1, 1.5, 2	$0, \pi$	4	200
2	440	0	...	0.5, 0.75, 1, 1.5, 2	...	4	200
3	440	50	6	0.5, 0.75, 1, 1.5, 2	$0, \pi$	4	200
4	440	100	4	0.5, 0.75, 1, 1.5, 2	$0, \pi$	4	200
5	440	100	6	0.5, 0.75, 1, 1.25, 1.5, 1.75, 2	$0, \pi/2, \pi, 3\pi/2$	20	2240
6	440	100	8	0.5, 0.75, 1, 1.5, 2	$0, \pi$	4	200
7	440	200	6	0.5, 0.75, 1, 1.5, 2	$0, \pi$	4	200
8	880	100	6	0.5, 0.75, 1, 1.5, 2	$0, \pi$	4	200
9	1500	100	6	0.5, 0.75, 1, 1.5, 2	$0, \pi$	4	200

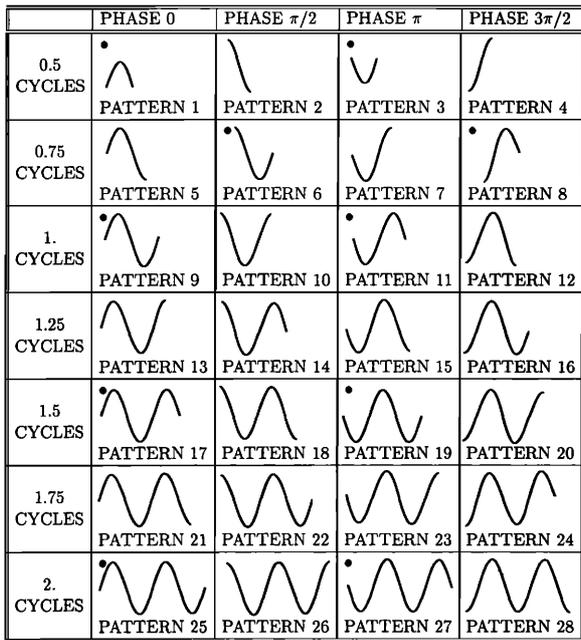


FIG. 3.  $F_0$  patterns of the 28 vibrato tones used in the main experiment. Initial phases: 1=0, 2= $\pi/2$ , 3= $\pi$ , 4= $3\pi/2$ . The patterns used for parametric experiments are emphasized using a black bullet.

because it has been tested for a larger group of subjects, and because the  $F_0$  pattern durations and phases have been studied in greater detail. A group of 20 musically educated subjects (group 1) was used for the main experiment. A smaller group of four selected subjects (group 2) was used for the preliminary and parametric experiments. All the subjects participated in several training and test sessions. All the recordings were made after some training. The group 2 subjects were all music graduates, and had much more training than the group 1 subjects.

Figure 3 shows the  $F_0$  patterns used for the main experiment. It must be emphasized that even the shortest tone may be found in a musical context, in ornaments or rapid staccato tones for instance. The selected patterns used for the parametric experiments are marked with bullets. We reduced the number of patterns and subjects in the parametric experiments, because of the large amount of time needed to test all the experimental conditions, and also because the main experiment demonstrated that this subset of selected patterns was the most informative. This subset is made of patterns ending with phase 0 and  $\pi$  only, and with durations 0.5, 0.75, 1, 1.5, 2 cycles.

A part of the preliminary and main experiments were completed in a first series of experimental sessions. The other part of the preliminary and main experiments, together with the parametric experiments were completed in a second series of experiments, approximately 8 month after the first series.

## II. RESULTS AND DISCUSSION

### A. Preliminary experiment

The results obtained for the preliminary experiment, using long tones, are similar to those reported in Shonle

TABLE III. Average settings (difference between the subjects responses and the center frequency) and standard deviations in cents for the main experiment, group 1.

Duration cycles	Initial phase 0	Initial phase $\pi/2$	Initial phase $\pi$	Initial phase $3\pi/2$
0.50	24.9 ± 9.7	-12.4 ± 9.8	-26.5 ± 7.7	14.8 ± 9.3
0.75	-0.8 ± 11.1	-17.8 ± 9.0	7.1 ± 14.5	19.2 ± 11.1
1.00	-12.8 ± 10.5	6.7 ± 13.2	12.4 ± 11.9	-4.4 ± 11.5
1.25	7.7 ± 8.4	12.5 ± 9.4	-5.5 ± 9.3	-11.4 ± 9.1
1.50	11.9 ± 9.7	2.0 ± 7.7	-12.2 ± 8.7	1.9 ± 9.2
1.75	3.2 ± 8.1	-8.0 ± 9.6	3.2 ± 8.6	7.6 ± 9.5
2.00	-4.7 ± 8.0	1.1 ± 7.6	3.7 ± 7.3	1.6 ± 6.9

and Horan (1980) although the experimental procedure is not exactly the same. We used synthetic vocal tones and a discrete adjustment task instead of pure tones and a continuous adjustment task. Table I shows that the results are comparable for both experiments. It is noticeable, for our results and for results reported by other researchers, that the standard deviations are quite high. This point will be discussed later. The preliminary experiment demonstrated the validity of our experimental procedure for pitch perception measurement. As a matter of fact, our preliminary experiment partly replicates previous experiments for long tones by Sundberg (1978), who carefully tested this type of method.

### B. Main experiment

Table III shows the results for the main experiment (group 1). Two different parameters are varied for each pattern in Fig. 3: The duration and the initial phase.

Figure 4 shows the means of Table III. The  $x$  axis represents the pattern duration. The  $y$  axis represents the difference in cents between the average pitch judgment and the center frequency. Patterns sharing the same final phase are connected by a similar line. The four different final phases that are present in our stimuli are therefore represented by four different lines. It is noticeable that these four lines converge toward the center frequency, without oscillations. It is clear from this figure that the pitch judgments were influenced most by the final phases (i.e., the shape of  $F_0$  patterns at the end of the tones). The seven patterns ending with phase  $\pi$  formed the top line (they were perceived higher for a same duration). The seven patterns ending with phase 0 formed the bottom line, and the patterns ending with phase  $\pi/2$  were perceived higher than those ending with phase  $3\pi/2$ , but lower than those ending with phase  $\pi$ . The first observation concerning the influence of phase, is that the pitch judgements are dependent on the final phase. The same data organized according to the initial phase instead of the final phase would have given four oscillating lines, indicating that the initial phases are not relevant parameters for explaining the pitch perceived.

The center frequency has been chosen as a reference in Fig. 4 because it is a fixed reference for all the stimuli. This choice can be questioned for stimuli 1 and 3 in Fig. 3, because in these cases the frequencies present in the stimuli are always either greater or lower than the center fre-

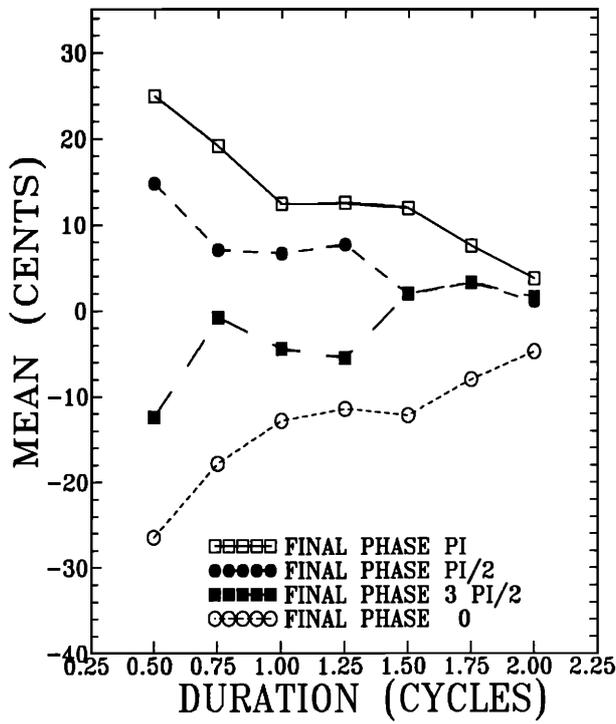


FIG. 4. Average settings for the main experiment. X axis: Pattern duration, in number of cycles. Y axis: Difference between the mean pitch perceived and center frequency. Patterns sharing the same final phase are connected by a line. center frequency: 440 Hz, vibrato extent: 100 cents, vibrato rate: 6 Hz.

quency. The center frequency equals the MBE for all the patterns excepted for patterns 1 and 3 in Fig. 3. The MBE for patterns 1 and 3 ( $\pm 25$  cents) is very close to the matchings obtained in the experiments. Nevertheless, for other patterns, one can notice a discrepancy between the pitch matchings and the MBE (which equals the center frequency).

An alternative presentation of the same data is proposed in Fig. 5. The LA of the stimuli is taken as reference frequency: The figure shows the difference between the settings and the LA pitch. For some patterns, the match between experimental data and LA pitch is very good (for instance patterns 1 and 3, 21 to 28, etc.), while for other patterns a discrepancy exists (patterns 2 and 4, 5 to 9, etc.). The LA pitch hypothesis does not take into account the direction of time: A pattern and its time-reversed image have the same LA (for instance patterns 9 and 11, 5 and 8, 2 and 4 etc., in Fig. 3). Comparing, for instance, pattern 7 (duration 0.75 cycles, final phase  $\pi/2$ ) and pattern 6 (duration 0.75 cycles, final phase 0), which have the same LA, one can notice that the settings for pattern 7 were about 16 cents higher than its LA, although the settings for pattern 6 were about 7 cents lower than the same LA. This can easily be explained if one consider the shape of the modulation patterns, postulating some weighting of the past in pitch perception.

The second observation is that the LA hypothesis is not sufficient to explain our experimental data. This is

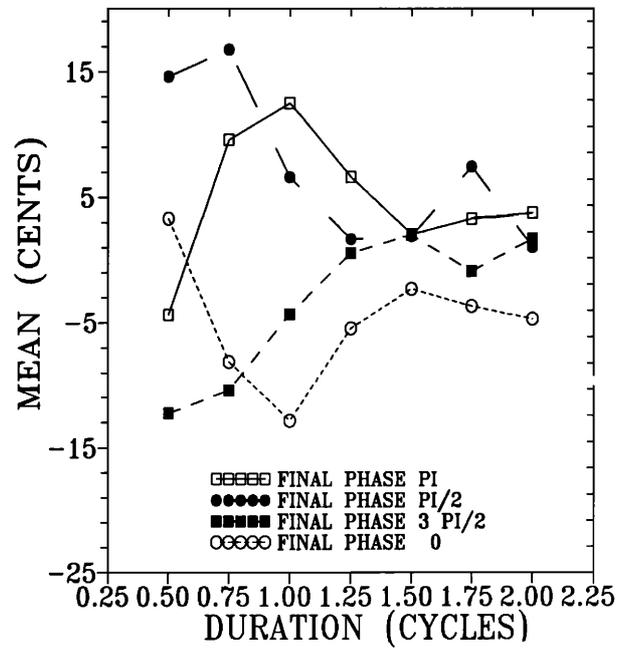


FIG. 5. Average settings for the main experiment. X axis: pattern duration, in number of cycles. Y axis: difference between the mean pitch perceived and the linear average of the stimuli. Patterns sharing the same final phase are connected by a line. Center frequency: 440 Hz, vibrato extent: 100 cents, vibrato rate: 6 Hz.

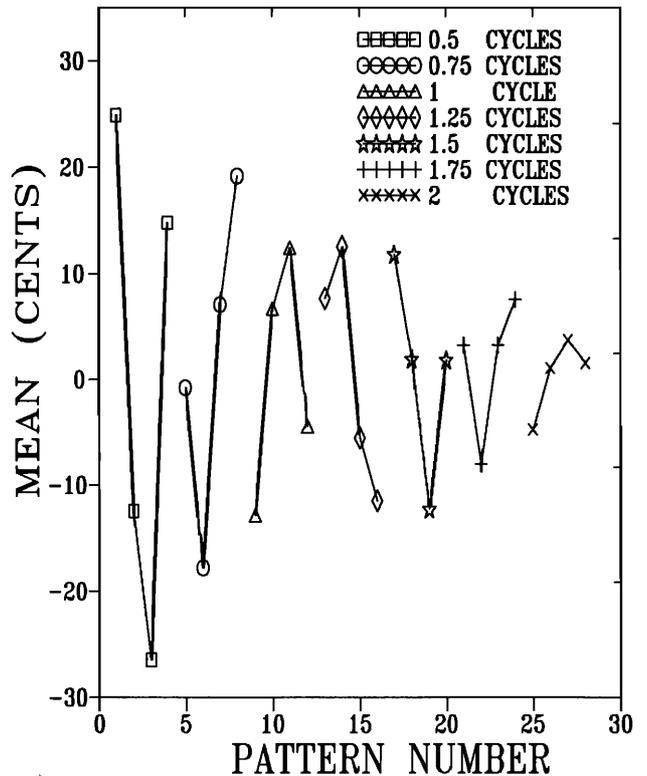


FIG. 6. Average settings for the main experiment. X axis: pattern number. Y axis: difference between the mean pitch perceived and center frequency. Patterns of the same duration are connected by a line. Center frequency: 440 Hz, vibrato extent: 100 cents, vibrato rate: 6 Hz.

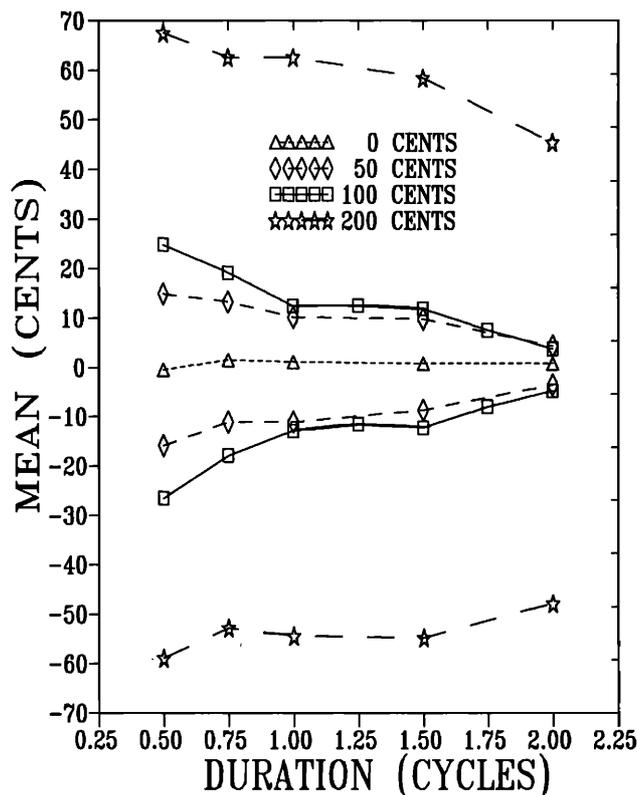


FIG. 7. Average settings for the parametric experiment on vibrato extent. X axis: pattern duration, in number of cycles. Y axis: difference between the mean pitch perceived and center frequency. Patterns sharing the same final phase are connected by a line. Center frequency: 440 Hz, vibrato rate: 6 Hz.

compatible with the first observation, because it indicates that some weighting of the past is necessary.

Figure 6 is another plot of the same data, where the X axis represents the pattern number, and where the Y axis represents the average pitch judgement. Patterns sharing the same duration are connected by a line. It is noticeable that the pairs of curves obtained for durations 0.5 and 1.5 cycles, 0.75 and 1.75 cycles, and 1 and 2 cycles have similar contours. These similar contours have decreasing amplitudes as duration increases. Again, this pseudoperiodicity

in pitch judgement is related to the periodicity of the patterns' final phase. The third remark, concerning the influence of duration, is that the pitch judgments are closer to the center frequency or the linear average as duration increases.

### C. Parametric experiment: Vibrato extent

Three parameters were studied in the parametric experiments as a function of duration and phase: Vibrato extent, center frequency, and vibrato rate. Figure 7 (using the same format as Fig. 4) displays the data reported in Table IV, obtained for vibrato extents of 0, 50, 100, and 200 cents. As in the presentation of the previous experiment, the data are organized according to the end of the tones. For extents 0, 50 and 100 cents, the lines clearly converge towards the center frequency.

Figure 8 displays the same data using the LA as reference frequency. The lines obtained for 50 and 100 cents conditions are similar.

For the 200 cents extent, all the tones were perceived as either significantly higher or lower than the center frequency almost independently of their duration. They were also perceived higher or lower compared to the LA, excepted for 0.5 cycles duration patterns (patterns 1 and 3).

This indicates that a separate perception of the high and low parts (extrema) of the  $F_0$  patterns appeared: The stimuli were perceived as two alternating tones rather than vibrato tones. Borrowing the terminology introduced in Nablèk *et 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 judgment, and the term "fusion" when the overall  $F_0$  pattern contributes to the pitch judgment. In the main experiment, for example, only the fusion case was observed, because the pitch perceived for patterns sharing the same end (same final arch) was dependent on the pattern durations, and therefore on the overall  $F_0$  patterns. In a musical context, fusion corresponds to true vibrato, and separation is rather related to trills or other melodic ornaments. Two observations emerge from this experiment: (1) If the vibrato extent is

TABLE IV. Parametric experiment: Vibrato extent. Average settings (difference between the subjects responses and the center frequency) and standard deviations in cents, center frequency 440 Hz, vibrato rate 6 Hz. ME means that these figures are those of the main experiment.

Number of Cycles	Initial phase (radians)	Vib. ex. 0 (cents)	Vib. ex. 50 (cents)	Vib. ex. 100 (cents)	Vib. ex. 200 (cents)
0.50	0	-0.39 ± 3.92	14.89 ± 4.25	ME	67.52 ± 6.52
0.50	$\pi$	...	-15.81 ± 6.58	ME	-58.83 ± 8.36
0.75	0	1.57 ± 4.02	-11.05 ± 7.20	ME	-52.73 ± 9.77
0.75	$\pi$	...	13.33 ± 5.04	ME	62.60 ± 9.59
1.00	0	1.18 ± 3.77	-11.05 ± 7.20	ME	-54.36 ± 12.08
1.00	$\pi$	...	10.20 ± 3.59	ME	62.60 ± 13.36
1.50	0	0.79 ± 3.40	9.81 ± 4.90	ME	58.42 ± 11.79
1.50	$\pi$	...	-8.68 ± 7.84	ME	-54.76 ± 12.75
2.00	0	0.79 ± 2.32	-3.54 ± 5.25	ME	-47.87 ± 22.10
2.00	$\pi$	...	4.32 ± 5.81	ME	45.43 ± 35.19

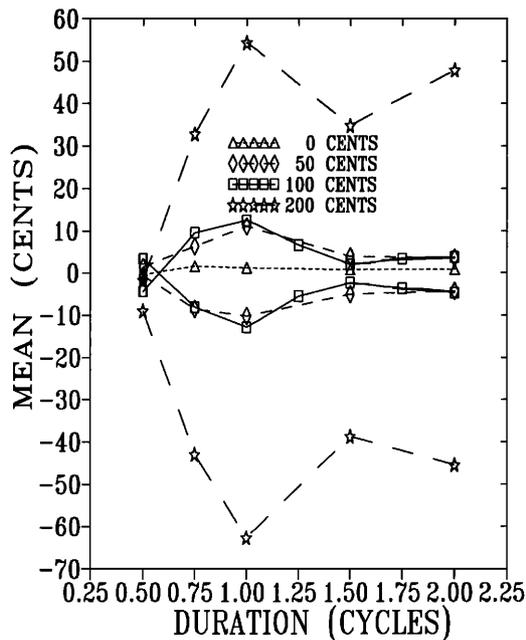


FIG. 8. Average settings for the parametric experiment on vibrato extent. *X* axis: pattern duration, in number of cycles. *Y* axis: difference between the mean pitch perceived and the linear average of the stimuli. Patterns sharing the same final phase are connected by a line. Center frequency: 440 Hz, vibrato rate: 6 Hz.

large, a separate perception of the upper and lower arches of the vibrato waveform occurs. This final arch alone contributes to pitch perception, the overall *F0* pattern has very little influence, if any, on the pitch perceived; (2) in the case of fusion, the final part of the *F0* pattern still seems the most important for pitch perception, but the overall *F0* pattern contributes to the pitch judgment.

#### D. Parametric experiment: Vibrato center frequency

Another parameter under study was the vibrato center frequency, as a function of duration and phase. Center frequencies of 220, 440, 880, and 1500 Hz were used. In Fig. 9 the data of Table V are organized according to the final phase of the patterns. This experiment indicates that the convergence toward the center frequency is almost independent of this center frequency. An inflection point (for a 1-cycle duration) is also noticeable for all the center frequencies. This inflection point will be explained later, in a model. A little difference between the different conditions is visible in Fig. 9: The 440-Hz condition settings are closer to the center frequency than the other condition settings. This difference can be related to the different groups of subjects: The 440-Hz condition was obtained for group 1, and other conditions for group 2. The subjects of group 2 were more trained than those of group 1. This difference seems to indicate that they focused more on the *F0* extrema.

#### E. Parametric experiment: Vibrato rate

The last parameter under study was the vibrato rate. Vibrato rates of 4, 6, and 8 Hz were used. In Fig. 10 the

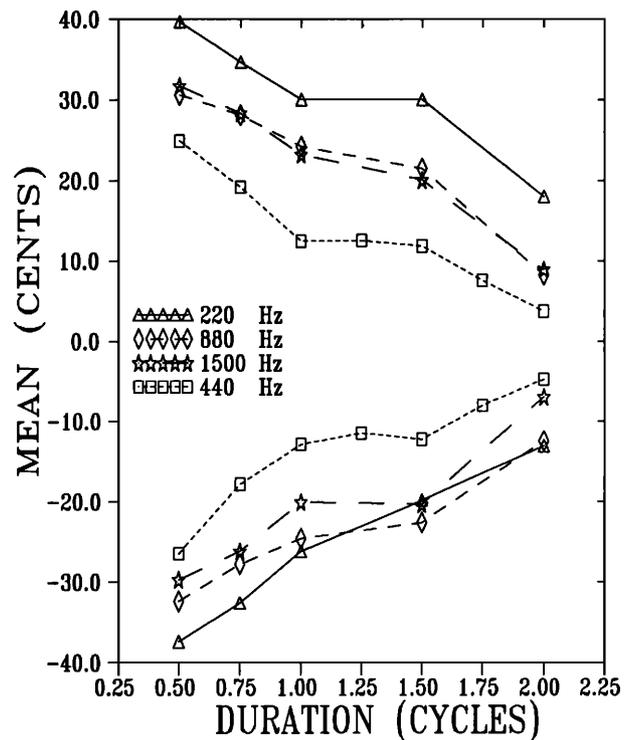


FIG. 9. Average settings for the parametric experiment on center frequency. *X* axis: pattern duration, in number of cycles. *Y* axis: difference between the mean pitch perceived and center frequency. Patterns sharing the same final phase are connected by a line. Vibrato extent: 100 cents, vibrato rate: 6 Hz.

data of Table VI are organized according to the final phase of the patterns. In this figure, the time axis represents duration expressed in seconds rather than a number of cycles. For this experiment, 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. This indicates a separate perception of the high and low parts (upper and lower arches of the sinusoidal waveforms) of the *F0* patterns. These short tones were perceived more as two alternating tones than as vibrato tones. It must be pointed out that the vibrato extent was only 100 cents, but that the vibrato cycles were rather long. For higher vibrato rates, one can observe that the lines converge toward the center frequency: It is the fusion case. Furthermore, the 8-Hz lines are converging faster than the 6-Hz lines. Figure 11 displays the same data using the LA as reference frequency. For the 8-Hz condition, one can notice that the LA and the pitch matchings are quite different for almost all the patterns (0.5 cycles patterns excepted).

It emerges from this last experiment that: (1) separation occurs for a slow vibrato rate (as was the case for a large extent); (2) in the case of fusion, the lines are converging faster for the higher vibrato rate. The difference between LA and pitch matchings is also higher.

These observations are compatible with the observation that the overall *F0* pattern contributes to pitch perception in case of fusion. For higher vibrato rates, the vibrato cycles are shorter and therefore the same pattern in

TABLE V. Parametric experiment: Center frequency. Average settings (difference between the subjects responses and the center frequency) and standard deviations in cents. Vibrato extent 100 cents, vibrato rate 6 Hz. ME means that these figures are those of the main experiment.

Number of cycles	Initial phase (radians)	N. freq. 220 (Hz)	N. freq. 440 (Hz)	N. freq. 880 (Hz)	N. freq. 1500 (Hz)
0.50	0	39.67 ± 4.90	ME	30.61 ± 4.17	31.74 ± 5.87
0.50	$\pi$	-37.39 ± 4.37	ME	-32.37 ± 4.86	-29.68 ± 7.70
0.75	0	-32.57 ± 7.44	ME	-27.76 ± 7.04	-26.17 ± 7.38
0.75	$\pi$	34.67 ± 4.65	ME	28.10 ± 5.64	28.33 ± 6.92
1.00	0	-26.17 ± 10.60	ME	-24.57 ± 6.67	-20.02 ± 7.81
1.00	$\pi$	30.03 ± 5.70	ME	24.22 ± 6.29	23.22 ± 6.14
1.50	0	30.03 ± 7.97	ME	21.51 ± 9.84	20.08 ± 6.92
1.50	$\pi$	-19.79 ± 9.48	ME	-22.57 ± 6.37	-20.32 ± 6.66
2.00	0	-13.03 ± 13.23	ME	-12.44 ± 6.38	-6.94 ± 5.60
2.00	$\pi$	18.01 ± 10.88	ME	8.24 ± 8.52	8.92 ± 9.35

terms of number of cycles is also shorter. This could explain a faster convergence.

### F. Interpretation of the results

The results obtained in the previous experiments indicate that the pitch judgments for short tones may be far from the mean between the extreme frequencies, as reported for long tones.

The pitch judgments corresponded to a time average of the overall  $F_0$  pattern (for perceptually fused patterns), or to a time average of the final arch of the  $F_0$  pattern (for

perceptually separated patterns). The hypothesis of a time average explains the periodicity observed in Fig. 6. It also explains the inflection points in Fig. 9 and the convergence toward the center frequency of the lines corresponding to the same final phase. But it is clear that a time-average model is not sufficient to explain our data, because for an entire number of vibrato cycles the time average equals the center frequency, although the pitch judgements obtained in these cases are not the center frequency. Another argument against the time-average model is the monotonic convergence toward the center frequency observed (in case of fusion). The comparison of our data with the linear average of the stimuli shows that a simple linear average is generally not able to represent the pitch perceived for short tones. The linear average is close to the pitch perceived only for the longer stimuli, or for the shorter stimuli (patterns 1 and 3). In these situations, both MBE and LA are also coinciding approximately. Therefore, it seems that the final part of the tone has a larger weight on the pitch judgment than the initial one. This could explain the converging monotonic lines obtained for data organized according to the end of the tones. A weighted time-average of the  $F_0$  pattern can be proposed to represent the experi-

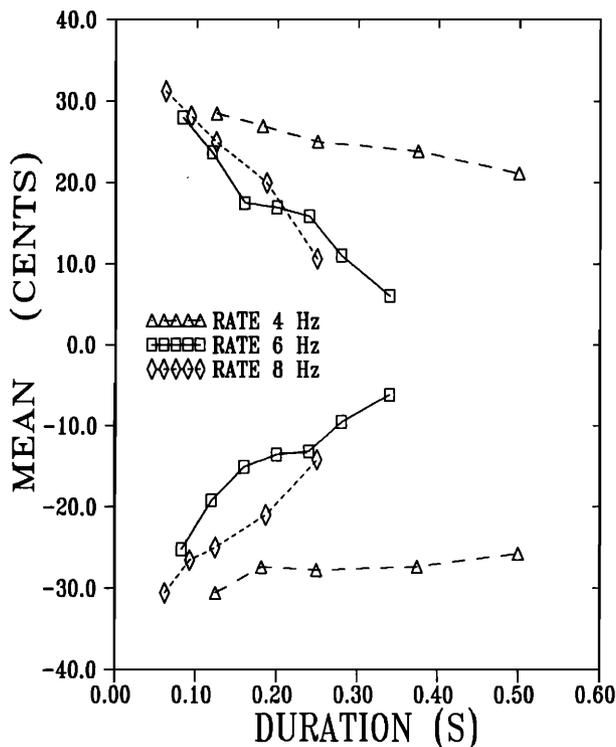


FIG. 10. Average settings for the parametric experiment on vibrato rate. X axis: pattern duration, in seconds. Y axis: difference between the mean pitch perceived and center frequency. Patterns sharing the same final phase are connected by a line. Vibrato extent: 100 cents, center frequency: 440 Hz.

TABLE VI. Parametric experiment: Vibrato rate. Average settings (difference between the subjects responses and the center frequency) and standard deviations in cents. Vibrato extent 100 cents, center frequency 440 Hz. ME means that these figures are those of the main experiment.

Number of cycles	Initial phase (radians)	V. freq. 4 (Hz)	V. freq. 6 (Hz)	V. freq. 8 (Hz)
0.50	0	28.49 ± 5.16	ME	31.19 ± 6.09
0.50	$\pi$	-30.56 ± 4.52	ME	-30.56 ± 5.73
0.75	0	-27.36 ± 5.25	ME	-26.56 ± 6.24
0.75	$\pi$	26.94 ± 8.79	ME	28.10 ± 7.20
1.00	0	-27.76 ± 5.83	ME	-24.97 ± 6.70
1.00	$\pi$	25.00 ± 7.71	ME	25.00 ± 8.48
1.50	0	23.84 ± 7.65	ME	19.95 ± 7.23
1.50	$\pi$	-27.36 ± 5.81	ME	-20.98 ± 6.70
2.00	0	-25.77 ± 6.52	ME	-14.22 ± 9.51
2.00	$\pi$	21.12 ± 9.67	ME	10.59 ± 9.05

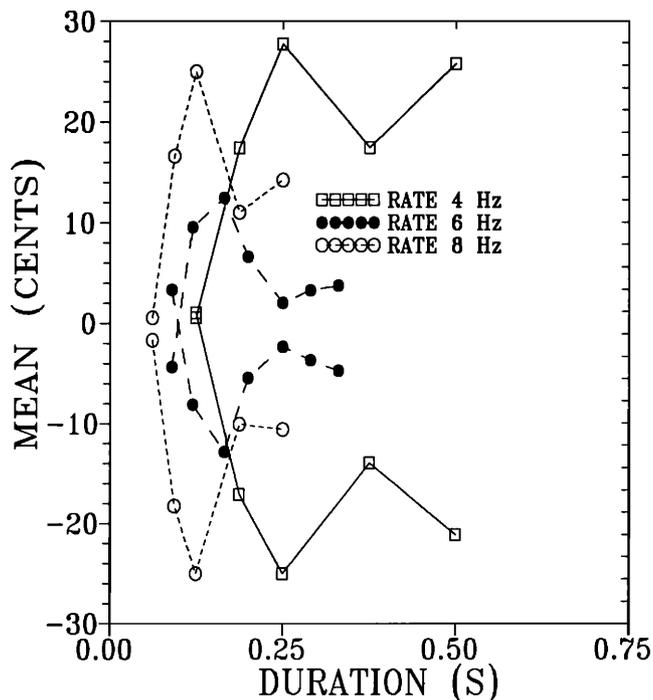


FIG. 11. Average settings for the parametric experiment on vibrato rate. X axis: pattern duration, in seconds. Y axis: difference between the mean pitch perceived and the linear average of the stimuli. Patterns sharing the same final phase are connected by a line. Vibrato extent: 100 cents, center frequency: 440 Hz.

mental data. Such a model will be discussed in detail in the next section.

### G. Constant pitch and gliding pitch

All the subjects noted that for several patterns it was possible to perceive a gliding pitch, rather than a constant pitch. Gliding pitch corresponds to a perceived tonal movement: The actual  $F_0$  pattern (the actual physical variation of the fundamental frequency) results in a psychologically gliding tone (or glissando). For constant pitch, the actual  $F_0$  pattern results in a single pitch. It must be emphasized that the gliding/constant pitch concept is linked to, but is distinct from, the fusion/separation concept. A “fused” response is a response taking into account the overall  $F_0$  pattern for pitch judgement, as opposed to a “separated” response taking into account only the final arch of the  $F_0$  pattern. A subject could report after an experiment that she/he perceived a gliding pitch, and nevertheless give a fused response.

Three situations were encountered: (1) A fused response and a stimulus perceived with a constant pitch; (2) a fused response and a stimulus perceived with a gliding pitch; (3) a separated response and a stimulus perceived with a gliding pitch. Situations 1 and 2 were present in the main experiment. Nevertheless the order of presentation (CV) or (VC) did not influence the results in this case (as discussed earlier). For our experiments, it seems that fusion/separation of the responses is the more relevant indication of perceptual integration. As a matter of fact,

TABLE VII. Maximum pitch changes for the stimuli, and glissando threshold according to 'tHart *et al.* ST=semitones. A factor of two distance from the threshold is also given.

Condition	Vibrato Rate Hz	Vibrato half period ms	Vibrato extent ST	Glissando rate ST/s	Glissando threshold ST/s
4	4	125	1	8	10.2[5.1–20.4]
3	6	83	0.5	6	23.2[11.6–46.4]
1 5 8 9	6	83	1	12	23.2[11.6–46.4]
7	6	83	2	24	23.2[11.6–46.4]
6	8	62	1	16	40.9[20.4–81.8]

stimuli with gliding or constant pitches gave rise to the same type of responses when fusion occurred, and to different types of responses only in case of separation.

The experiment on the presentation order was made when designing the test methodology, and we tested the presentation order only in the fusion case. It is clear that in the separation situation the presentation order could influence the responses: As glissandi between two alternating tones are perceived, the subjects could focus on the high tone or the low tone, depending on the experimental instructions.

### H. Fusion/separation and the glissando threshold

As the fundamental frequency changes continuously in vibrato tones, it is important to know in which conditions these changes are audible, and in which conditions they go unnoticed. This problem is related to the glissando threshold. This threshold is also called the absolute threshold of pitch change, and is expressed in frequency unit per unit of time.

Table VII shows the maximum  $F_0$  changes that are present in our data: The maximum  $F_0$  change is obtained as the vibrato extent divided by half the vibrato period. This maximum  $F_0$  change is actually reached, for instance, for patterns 2 and 4 in Fig. 3.

Psychoacoustic and psychophysical data on the glissando threshold have been obtained by Klatt (1973), Pollack (1968), Rossi (1971, 1978), Schouten (1985), and Sergeant and Harris (1962). No reported experiment exactly matches our experimental conditions (regarding tone duration, frequency change, and center frequency). It is therefore difficult to extract useful information from this large body of data, and the reported results must be interpolated for our purpose. In fact, an interpolation procedure and a unified view of this problem was presented by 'tHart *et al.* (1990). They proposed the semitone/second as the best unit for the glissando threshold, because it appeared that the threshold is almost independent on the center frequency using a relative frequency scale. They studied the distribution of the glissando thresholds published in the literature. They showed that the glissando thresholds were distributed around a curve  $G_{tr}$  (expressed in semitones/s) which approximately satisfies the equation:

$$G_{tr} = 0.16/T^2, \quad (3)$$

where  $T$  is the duration of the tone. If Eq. (3) is plotted using a double (natural) logarithmic scale, it becomes approximately a straight line, in our domain of variation:  $T \in [0.06 \text{ s}, 0.130 \text{ s}]$

$$\log(G_{tr}) = -2.00 \times \log(T) - 1.83. \quad (4)$$

'tHart *et al.* reported that more than 75% of the data in the literature lie within a distance of a factor of 2 from Eq. (3), i.e., within the interval  $[\log(G_{tr}) - \log(2), \log(G_{tr}) + \log(2)]$ , in the double logarithmic scale. In Table VII, we reported the pitch changes encountered in our experiment, compared to the interpolated thresholds, and the 75% interval as discussed by 'tHart *et al.*

The fusion/separation situations observed in the parametric experiment are in good agreement with the glissando threshold. The glissando rates in case of separation (lines one and four of Table VII) are in the 75% interval around the glissando threshold. In case of fusion, the glissando rates are all below the threshold. For the main experiment (line three in Table VII), the glissando rate is close to the lower boundary of the 75% interval. This means that the pitch change might have been noticeable for this experiment (and as a matter of fact it was actually reported by some of the subjects). Nevertheless, the results demonstrated fused responses for this experiment: In this case the overall  $F_0$  change influenced the pitch judgment. When the pitch change was clearly below the glissando threshold, however, no pitch change at all was perceived (e.g., for half a cycle of 8-Hz vibrato rate or 50 cents vibrato extent stimuli).

A somewhat similar fusion/separation situation (the trill threshold) has been introduced for long-duration tones by Miller and Heise (1950) and later reconsidered by Shonle and Horan (1976). It has been shown that fusion tends to fail when the vibrato rate comes below 5 Hz, and when the vibrato extent comes above 200 cents, in the 500-Hz frequency region. Although the work on the trill threshold concerned only long tones, the results seems in good agreement with our findings for short tones, and with the glissando threshold.

### I. Standard deviations

In spite of the reported ambiguity of some patterns, when a glissando was perceived, the standard deviations obtained were generally comparable to those reported for long tones. Cardozo (1965) found that the difference limen was about a factor 1.7 larger than the corresponding standard deviation for a method of adjustment. This is not the case for our experiments. It was not the case for other experiments on vibrato; for instance Shonle and Horan (1980, p. 251) reported that their standard deviations were larger than those predicted by Cardozo. The frequency-difference limens for short tones are duration dependant [see Moore (1973)], but are still lower than those computed on the basis of our experimental results. In our opinion, this discrepancy might be explained by the particular  $F_0$  patterns used: In frequency-difference limen experi-

ments, the  $F_0$  patterns are flat, and it is likely that frequency modulation introduced by vibrato degrades the difference limens. For instance, Jesteadt and Sims (1975) compared performances for the same group of subjects with pulsed sinusoid and frequency modulated tones, and found the difference limens for frequency-modulated tones to be consistently larger.

It must be emphasized that the standard deviations were of the same order of magnitude in almost all the experiments. It is therefore difficult to interpret consistently the differences in standard deviation.

The only cases of larger standard deviations were in the parametric experiment, for the 200 cents vibrato extent condition (condition 8), and for the 1 and 2 cycles pattern durations. These stimuli are in the separation situation. They have an integer number of cycles. It seems that they are the most ambiguous stimuli. This reported ambiguity might be explained by the conflict between the separate perception of the last arch of the  $F_0$  pattern, and the perception of the center frequency. On the one hand, as there is separation, a first candidate for pitch perception is the time average of the last arch of the vibrato waveform. But on the other hand, as there is an integer number of vibrato cycles, the  $F_0$  pattern begins and ends on the center frequency, and is symmetric (one maximum and one minimum, or two maxima and two minima). This might have encouraged the subjects to choose the center frequency as the best candidate for pitch perception. The results show that the average response agrees with the first alternative (time average of the last arch), but the large standard deviation might be an indication of ambiguity.

### J. Absolute pitch

The question of absolute pitch may also be raised for our experiments. Categorical pitch judgments, related to the tempered scale, could have been reported by some subjects, instead of an averaging of the  $F_0$  patterns. As a matter of fact, this type of judgment was obtained for a subject with a pure form of absolute pitch. This subject is also a professional piano performer, specialized in accompaniment of singers. She is therefore particularly used to  $F_0$  patterns of real singers. It is not clear whether her categorical responses are due to absolute pitch, or whether they are a product of her professional interest in singers' performances. Moreover, other subjects with absolute pitch gave averaged judgments rather than categorical ones. With the exception of one subject in group 1, the results of subjects with absolute pitch are similar to the results of subjects without absolute pitch.

## III. NUMERICAL MODEL

### A. Equation of the model

We have seen in the previous sections that the final part of the tone had a larger weight on the pitch judgment than the initial one. The experimental results also suggested that the  $F_0$  patterns were time averaged, because of the convergence toward the center frequency for long tones, at least in case of fusion.

Let us consider first the case of fusion. A quantitative model for such a process may be a time average of the  $F0$  pattern viewed through a data window. We shall refer to this model as the weighted time average (WTA) model. 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. A simple time average is clearly not sufficient, as discussed earlier, but an exponential memory function is also not sufficient, because only the recent past is taken into account using such a weighting function. A combination of the two is required (raised exponential memory function). Let  $p(t)$  denote the pitch perceived at time  $t$ ,  $f$  the time-varying  $F0$  function beginning at time 0, and let  $\alpha$ ,  $\beta$  be two constants:

$$p(t) = \frac{\int_0^t (e^{-\alpha(t-\tau)} + \beta) f(\tau) d\tau}{\int_0^t (e^{-\alpha(t-\tau)} + \beta) d\tau}. \quad (5)$$

The meaning of Eq. (5) is twofold: The constant  $\beta$  accounts for the time averaging, and the constant  $\alpha$  accounts for the weighting of the past.

In the case of separation, the excitation patterns due to successive vibrato extrema become more separated in time or in frequency, and these extrema are perceived as two independent auditory events. It is necessary to introduce in the model a time-frequency threshold related to the glissando threshold.

The glissando rate for a given vibrato extent (VE) and a given vibrato rate (VR) can be computed as the product of the extent and two times the rate. It is reasonable to assume that if the glissando rate for a given condition is larger than the glissando threshold for the same condition, the amount of constant time averaging represented by  $\beta$  is reduced. It does mean that in the separation case, the pitch judgment does not take into account the distant past. The numerical model is therefore given by Eq. (5), with  $\beta=0$  if  $VE \times 2VR > G_{tr}$ , where  $G_{tr}$  is given by Eq. (3).

## B. Parameters estimation and results

Only two free parameters  $\alpha$  and  $\beta$  have to be estimated. The model must be able to represent the data obtained both in the main experiment and in the parametric experiment. To estimate the model parameters, it is classical to minimize the distance between the model response and the experimental data. Several sets of data can be taken as reference in the minimization process. Due to the statistical nature of our results, we chose to minimize the root-mean-square distance between the results obtained with the two-parameters model and the data of the main experiment (Table III). This two-dimensional rms minimization gives the optimal parameters  $\alpha = -22$  and  $\beta = 0.20$ . The parameter  $\beta$  can be interpreted as the amount of long-term time average (here  $\beta = 20\%$ ). The parameter  $\alpha$  represents the speed of decay of the exponential function. This weighting process can be considered as a low-pass filtering process, taking the exponential function in Eq. (5) as the impulse response of the system. In this case, one can estimate the bandwidth  $B$  of the low-pass filter:  $B = \alpha/\pi = 7.00$  Hz, and its time constant is  $1/B = 0.14$  s.

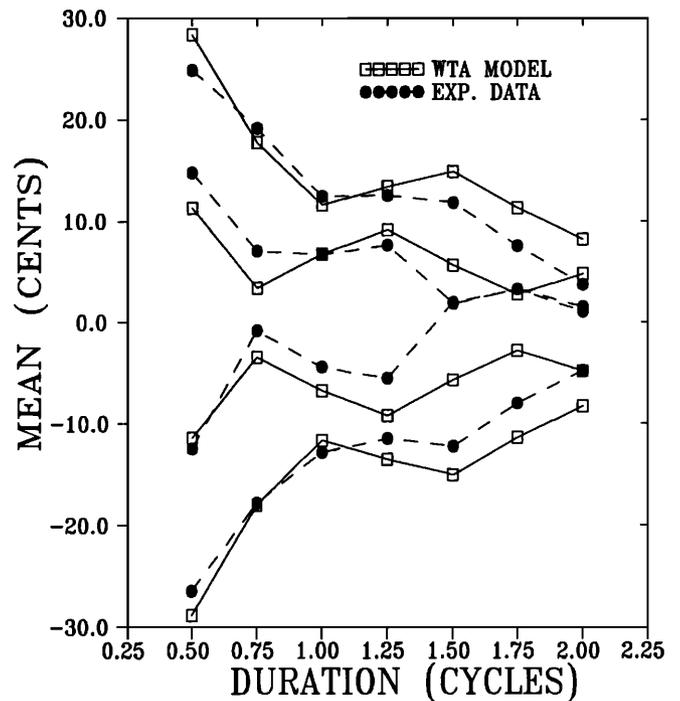


FIG. 12. Weighted time average model of pitch perception. Data of the main experiment and result of the numerical model.

The figure of 7 Hz (or 0.14 s) was extensively discussed by Stevens and Davis (1938, pp. 222–223, 237–239, 242), in the context of vibrato. They reported that the gliding character of vibrato tones is lost for a vibrato rate faster than 7 Hz: “But then, as the rate is increased, although the maximum continues to move back and forth, the movement no longer appears as a change of pitch.” They also compared this result with previous work of Békézy: “The figure of 0.14 sec reminds us that Békézy reported that the persistence of an auditory sensation lasts about this length of time.” For beats also, Stevens and Davis reported that: “At the rate of about 6 or 7 beats per second, ..., the smooth rise and fall in loudness vanishes and each beat appears as a single impulse. ... Beyond this critical rate, we are left, in both instances (i.e., vibrato and beats), with a tone having an intermittent, throbbing character.” Therefore, it seems that the optimal integration time constant that was obtained for our data, can be consistently related to a psychophysical time threshold, the auditory persistence.

Figure 12 displays the results obtained with this model, for the optimal parameters, and the conditions of the main experiment, using the center frequency as reference. Figure 13 displays the discrepancy between experimental data and the WTA pitch. This figure can be compared to the same figure (Fig. 5) using the LA model. Figure 14 compares the LA model and the experimental data of the main experiment, using also the center frequency as reference. It is clear that the match between model and experimental data is better for the WTA model

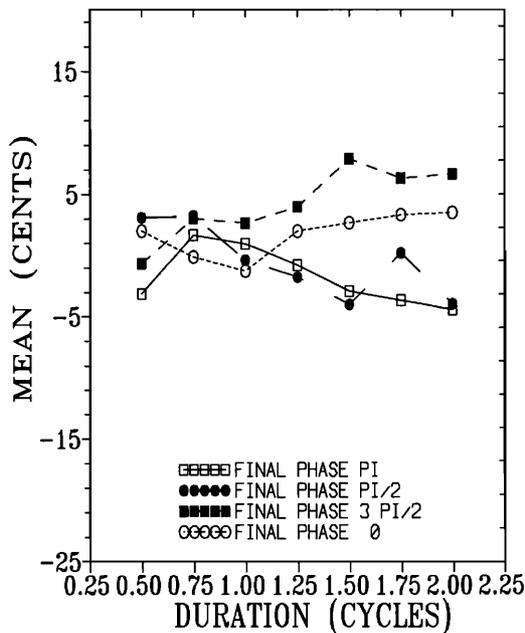


FIG. 13. Weighted time average model of pitch perception. X axis: pattern duration, in number of cycles. Y axis: difference between the mean pitch perceived (main experiment) and the result of the numerical model.

than for the LA model, particularly for the 1 or 2 cycles stimuli.

Figures 15 and 16 show the application of the numerical model to the parametric experiments. This numerical model seems well suited to both fused and separated vibrato judgments. It must be pointed out that the optimal model parameters obtained for a first set of experimental

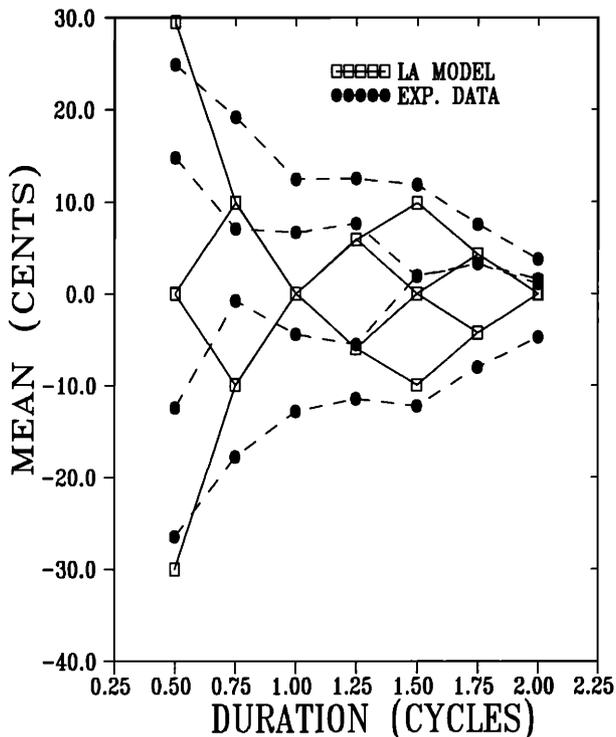


FIG. 14. Linear average model of pitch perception. Data of the main experiment and linear average of the stimuli.

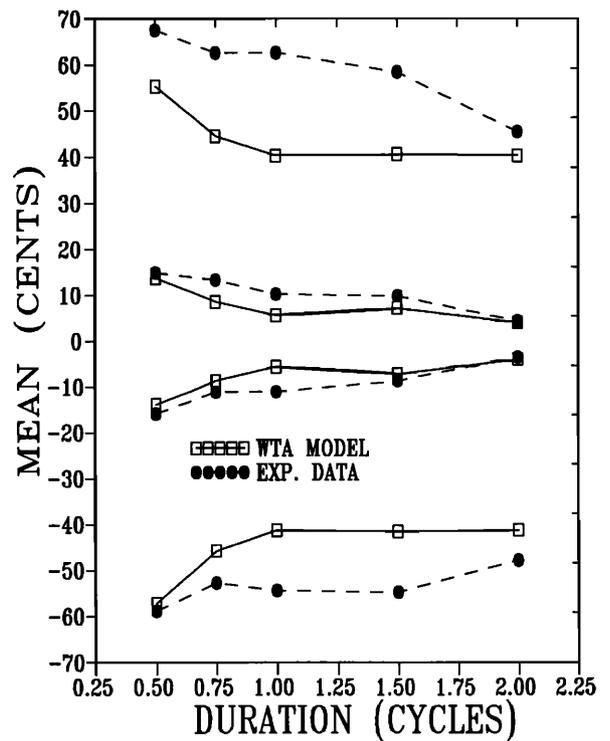


FIG. 15. Weighted time average model of pitch perception. Data of the parametric experiment on vibrato extent and result of the numerical model.

data (the main experiment) represent quite well a second set of the experimental data (parametric experiment).

#### IV. GENERAL DISCUSSION AND SUMMARY

Our results demonstrate that for short tones the pitch does not correspond to the mean frequency between the peaks of  $F_0$ , as seems to be the case for long tones. An alternative hypothesis for long tones was proposed by Sundberg (1978): The linear average hypothesis. Shonle and Horan (1980) also proposed the LA hypothesis in a subsidiary experiment, using asymmetrical vibrato waveforms: A linear rise from the lowest value to the highest value for one quarter of the vibrato cycle, a steady portion for half the period, and a linear fall for the remaining quarter period ("flat-top" condition, the "flat-bottom" condition was similar, except that the frequency remained steady at the lowest value for half a period). The results obtained indicated an averaging of all frequencies present (with some emphasis on the steady portions), rather than a mean between the extreme frequencies. It is difficult to compare this subsidiary experiment with ours, because no indications of phase or duration are available. There is also some discrepancy between the linear average of the stimuli and our experimental data. Nevertheless, the same type of mechanism could explain pitch perception for both short-duration and long-duration tones. This mechanism could be a weighted time average of the  $F_0$  pattern.

A previous model of weighted time average was presented by Feth and his colleagues in a series of studies on pitch of two-component and narrow-band signals. Feth

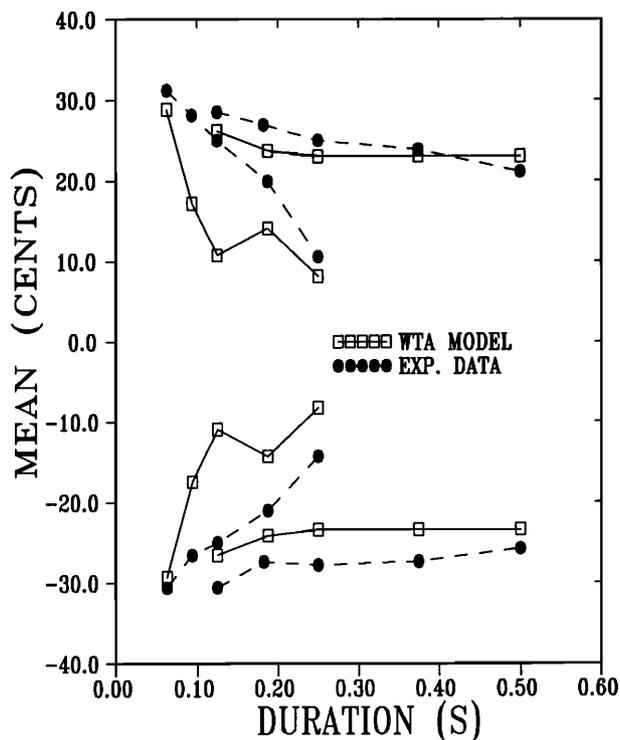


FIG. 16. Weighted time average model of pitch perception. Data of the parametric experiment on vibrato rate and result of the numerical model.

(1974) has reported that listeners are able to discriminate between two-tone complexes on the basis of their pitch [see for instance the discussion in Feth and O'Malley (1977), p. 941, and Stover and Feth (1983), p. 1706]. The pitch differences for such signals were consistently related to differences in the envelope-weighted averages of their instantaneous frequencies. This model is analogous to our time average of  $F$  model but does an envelope-weighted time average on the instantaneous frequency rather than fundamental frequency.

The pitch vibrato also occurs in sustained tones of various instruments. Violin vibrato tones were studied by Fletcher and Sanders (1967), and Seashore (1938). It seems (1) that the vibrato is present in practically all tones, open strings excepted; (2) that for long tones, the mean pitch of the tones tends to coincide with the pitch indicated in the printed score; (3) that the stopping finger moves both above and below the target pitch; (4) that the movement responsible for the pitch vibrato is initiated most frequently with the movement toward the bridge, and the final vibrato movement is most frequently toward the scroll. This means that for long tones as well as for short tones the  $F_0$  patterns are most frequently starting at the target frequency with phase 0 and ending at the target frequency with phase  $\pi$ . This is a somewhat simplified situation, compared to vocal vibrato. Our model of pitch perception seems in agreement with the observation that the target frequency (which is reached at the end of tones) is actually the perceived pitch, because the end of tones plays a particular role in this model. It is reasonable to think that our model works not only for vocal vibrato

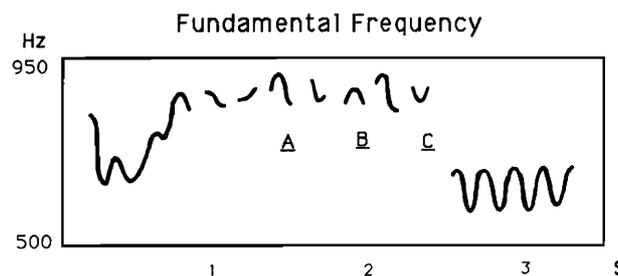


FIG. 17. Fundamental frequency pattern of a short sequence extracted from Mozart's Magic Flute: Aria of the Queen of the Night "Hell's vengeance seethes in my heart," sung by Beverly Hoch, CD EMI 7 54287 1/2/4/.

tones but also for instrumental vibrato tones.

Let us consider again Fig. 1. Figure 17 displays the fundamental frequency patterns of the musical sequence in Fig. 1. In this example many of the  $F_0$  patterns of Fig. 3 may be found [pattern 5 above letter (A) in Fig. 17, pattern 1 above (B), pattern 3 above (C) ...]. Our results are able to explain the  $F_0$  movements that are apparent in Fig. 17. For instance, the  $F_0$  pattern No. 1 (B) has a center frequency lower than the  $F_0$  pattern No. 3 (C), in order to be perceived with the same pitch.

In summary, we found that for short-duration vibrato tones, the pitch perceived corresponds to a weighted time average of the  $F_0$  pattern. Therefore, we have the following conclusions:

- (1) The overall  $F_0$  pattern influences pitch perception. The patterns duration and phase are important parameters for pitch perception.
- (2) Pitch depends most on the end of the tone, and it converges toward the mean between the  $F_0$  peaks (or toward the linear average of  $F_0$  over one vibrato period), as duration is increasing, at least for symmetric vibrato waveforms.
- (3) Separation/fusion of pitch perception depends on the vibrato extent and rate. This may be explained using the glissando threshold.

A simple numerical model which is able to represent our experimental data was proposed. This model consists of a weighted time average of the  $F_0$  pattern. The weighting function is a raised decaying exponential function. The model incorporates also a threshold (glissando threshold), in order to account for the separation/fusion phenomenon observed for large vibrato extents or slow vibrato rates.

## ACKNOWLEDGMENTS

The authors are indebted to the subjects for their kind help in the course of this research (particularly Sophie Grau, Frédéric Beaugendre, and Gaël Richard). We would like to thank Xavier Boutillon, Jean-Pierre Rossi, Jean-François Jodoin, Jean-Sylvain Liénard, and three reviewers for helpful comments on earlier versions of this paper.

<sup>1</sup>Other presentation conditions [(CV) or (VCVC) instead of (VC)] were tried, for a limited number of subjects. The differences between results obtained with different presentation conditions were not statistically sig-

nificant. This demonstrated that no masking effect resulting from the presentation method was noticeable. As no significant differences resulted, we chose the same (*VC*) presentation condition for all the experiments. Using the same arguments, if the 300-ms silent interval duration had been critical, some variation might have resulted between the (*CV*) and (*VC*) presentation conditions. It must be noted that the different test presentations were tried for a set of parameters corresponding to the main experiment (vibrato center frequency 440 Hz, vibrato extent 100 cents, vibrato rate 6 Hz). It is not certain that the (*CV*) and (*VC*) presentations are equivalent for a larger extent or a slower rate, as it will be discussed later.

<sup>2</sup>We used the frequency-difference limens reported in the classic paper by Shower and Biddulph (1931). These data are probably an upper boundary for the estimates of frequency-difference limens reported by other authors [see for instance a review in Wier *et al.* (1976)]. This choice was made for the sake of economy, as the number of stimuli was directly proportional to the frequency step. The main experiment was also performed, for a limited number of subjects, using the frequency-difference limens reported by Moore for short tones (Moore, 1973). These data are probably a lower boundary. No significant differences were obtained, either in terms of average judgment or in terms of standard deviation, and we feel therefore entitled to use Shower and Biddulph data.

<sup>3</sup>The absolute duration depends on both the number of vibrato cycles and the vibrato frequency.

d'Alessandro, C., and Castellengo, M. (1991). "Perception of vocal pitch vibrato in short tones," *J. Acoust. Soc. Am.* **89**, 1987.

d'Alessandro, C., and Castellengo, M. (1992). "A parametric study on pitch of short vocal vibrato tones," *J. Acoust. Soc. Am.* **91**, 2434.

Castellengo, M., Richard, G., and d'Alessandro, C. (1989), "Study of vocal pitch vibrato perception using synthesis," in *Proceedings of the 13th International Congress on Acoustics* (Sava Centar, Belgrade), pp. 113–116.

Cardozo, B. L. (1965). "Adjusting the Method of Adjustment: SD vs DL," *J. Acoust. Soc. Am.* **37**, 786–792.

Feth, L. L. (1974). "Frequency discrimination of complex periodic tones," *Percept. Psychophys.* **15**, 375–779.

Feth, L. L., and O'Malley, H. (1977). "Two-tone auditory spectral resolution," *J. Acoust. Soc. Am.* **62**, 940–947.

Fletcher, H., and Sanders, L. C. (1967). "Quality of violin vibrato tones," *J. Acoust. Soc. Am.* **41**, 1534–1544.

Klatt, D. H. (1973). "Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception," *J. Acoust. Soc. Am.* **53**, 8–16.

Jesteadt, W., and Sims, S. L. (1975). "Decision processes in frequency discrimination," *J. Acoust. Soc. Am.* **57**, 1161–1168.

Miller, G. A., and Heise, G. A. (1950). "The trill threshold," *J. Acoust. Soc. Am.* **22**, 637–638.

Moore, B. C. J. (1973). "Frequency difference limens for short-duration tones," *J. Acoust. Soc. Am.* **54**, 610–619.

Nabèlèk, I. V., Nabèlèk, A. K., and Hirsh, I. J. (1970). "Pitch of tone bursts of changing frequency," *J. Acoust. Soc. Am.* **48**, 536–553.

Pollack, I. (1968). "Detection of rate of change of auditory frequency," *J. Exp. Psychol.* **77**, 535–541.

Rossi, M. (1971). "Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole," *Phonetica* **23**, 1–33.

Rossi, M. (1978). "La perception des glissando descendants dans les contours prosodiques," *Phonetica* **35**, 11–40.

Schouten, H. E. M. (1985). "Identification and discrimination of sweep tones," *Percept. Psychophys.* **37**, 369–376.

Seashore, C. E. (1938). *The Psychology of Music* (McGraw-Hill, reprinted by Dover, New York, 1967).

Sergeant, R. L., and Harris, J. D. (1962). "Sensitivity to unidirectional frequency modulation," *J. Acoust. Soc. Am.* **34**, 1625–1628.

Shonle, J. I., and Horan, K. E. (1976). "Trill threshold revisited," *J. Acoust. Soc. Am.* **59**, 469–471.

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

Shower, E. G., and Biddulph, R. (1931). "Differential pitch sensitivity of the ear," *J. Acoust. Soc. Am.* **3**, 275–287.

Stevens, S. S., and Davis, H. (1938). *Hearing. Its Psychology and Physiology* (AIP, New York, 1983), 2nd printing.

Stover, L. J., and Feth, L. L. (1983), "Pitch of narrow-band signals," *J. Acoust. Soc. Am.* **73**, 1701–1706.

Sundberg, J. (1978). "Effects of the vibrato and the 'singing formant' on pitch," *Musica Slovaca* **VI** 51–69.

'tHart, J., Collier, R., and Cohen, A. (1990). "A perceptual study of intonation" (Cambridge U.P., UK), pp. 29–33.

Wier, C. C., Jesteadt, W., and Green, D. M. (1976). "A comparison of method-of-adjustment and forced-choice procedures in frequency discrimination," *Percept. Psychophys.* **19**, 75–79.