The pitch of short-duration fundamental frequency glissandos

Christophe d’Alessandro, Sophie Rosset, and Jean-Pierre Rossi
LIMSI–CNRS, BP 133 F-91403 Orsay, France

(Received 9 October 1997; accepted for publication 7 July 1998)

Pitch perception for short-duration fundamental frequency (F0) glissandos was studied. In the first part, new measurements using the method of adjustment are reported. Stimuli were F0 glissandos centered at 220 Hz. The parameters under study were: F0 glissando extents (0, 0.8, 1.5, 3, 6, and 12 semitones, i.e., 0, 10.17, 18.74, 38.17, 76.63, and 155.56 Hz), F0 glissando durations (50, 100, 200, and 300 ms), F0 glissando directions (rising or falling), and the extremity of F0 glissandos matched (beginning or end). In the second part, the main results are discussed: (1) perception seems to correspond to an average of the frequencies present in the vicinity of the extremity matched; (2) the higher extremities of the glissando seem more important; (3) adjustments at the end are closer to the extremities than adjustments at the beginning. In the third part, numerical models accounting for the experimental data are proposed: a time-average model and a weighted time-average model. Optimal parameters for these models are derived. The weighted time-average model achieves a 94% accurate prediction rate for the experimental data. The numerical model is successful in predicting the pitch of short-duration F0 glissandos. © 1998 Acoustical Society of America.

PACS numbers: 43.66.Hg, 43.75.Bc [DWG]

INTRODUCTION

In natural audio signals like speech or music, fundamental frequency (F0) is almost never constant. The main psychological correlate of F0 is pitch. It is well known that the physical aspect (F0) and the psychological aspect (pitch) can differ significantly, particularly for sounds whose fundamental frequency changes. The aim of the present article is to report new measurements and to propose a numerical model for pitch perception of short-duration F0 glissandos. Short-duration F0 glissandos are thought of as prototypes or building blocks for the F0 contours that can be encountered in many natural signals, like notes in music or syllables in speech.

Several theories accounting for human instantaneous pitch detection are available, together with the corresponding pitch detection algorithms [see Hess (1983) or Hermes (1993) for reviews]. These theories and techniques focus on instantaneous, or at least short-term, attributes of the acoustic signal. Only a few periods of signals are considered in order to compute the time-varying instantaneous fundamental frequency. However, the link between instantaneous F0 contours and pitch perception for larger units, like F0 glissandos, is by no means straightforward. It is therefore important to distinguish between instantaneous F0, which is a relatively well-known problem, and pitch perception for short-duration F0 glissandos.

Perceptual thresholds for predicting under what conditions F0 changes are audible have long been studied. A first perceptual threshold is the glissando threshold, or absolute threshold of pitch change. It is expressed in frequency unit per unit of time. Below the threshold, a F0 glissando is perceived with constant pitch. Above the threshold, a pitch glissando is perceived. Data on the glissando threshold for speechlike sounds have been obtained by Klatt (1973), Rossi1 (1971, 1978a), and Schouten (1985), following the pioneering work of Sergeant and Harris (1962) for pure tones. A unified view of this problem was presented by ′t Hart (1976) and ′t Hart et al. (1990), who conclude that the threshold can be approached by $G_f = 0.16/T^2$, where $G_f$ is the glissando threshold (in semitones/s) and $T$ the tone duration. When a glissando is above the glissando threshold, a separate perception of the high and low extremities of the F0 patterns appears. Nabelek et al. (1970) first studied this effect in a work on the pitch for tone bursts of changing frequency. They coined the term “separation” for when only the final part of the burst contributes to the pitch judgment, and the term “fusion” for when the overall F0 pattern contributes to the pitch judgment. Thresholds for separation and fusion are consistent with the glissando threshold.

Thresholds can predict whether a given F0 glissando (with a certain duration and a certain extent) will be perceived as a constant pitch tone or as a pitch glissando. However, the data on perceptual thresholds for pitch changes do not address the problem of what pitch is actually perceived, i.e., what the subjective pitch value is at the beginning and end of the F0 glissando. The first aim of this study is to give explicit measurements of the pitch perceived at the beginning and end of F0 glissandos. One of the few studies addressing this problem can be found in Rossi (1971, 1978a), where natural and synthetic vowels were used. Unfortunately, the experimental conditions used by Rossi were not likely to give accurate results, because of the choice of subjects (a group of students in phonetics whose individual capabilities in pitch discrimination were not tested) and because of the stimulus presentation method (loudspeakers in a classroom). Therefore, new and more accurate measurements were needed.

Another issue which needs to be reinvestigated is whether there are asymmetries in the perception of superliminar rising and falling F0 glissandos. Several articles
have noted such asymmetries, but the results have not always been consistent. Hombert (1975) and Rump and Hermes (1996) found that the high register has more weight in intonation perception. Brady (1961) and Wieringen and Pols (1995) reported that for pure tone glides, the end of the tone has more perceptual weight than the beginning. For pure tones and complex tones, Demany and McAnally (1994) and Demany and Clément (1995, 1997), found an asymmetry between peaks and troughs for large frequency modulations. However, Rossi (1978a) reported that no differences occurred in the perception of rising and falling glissandos. In this article, we compare the perception of rises and falls over a wide range of conditions.

The final purpose of this work is to propose a numerical model of pitch perception for short-duration $F_0$ glissandos. A first model addressing this problem was proposed by Rossi (1971, 1978a), using natural and synthetic vowels. Rossi postulated that for vowels with $F_0$ glissandos (either rising or falling), the pitch perceived at the end of vowels corresponds to an averaging of $F_0$ toward the end of the glissando. This is expressed in Rossi’s so-called “2/3 rule,” which can be stated as follows:

For dynamic tones in a vowel, the pitch perceived corresponds to a point between the second and the third third of the vowel.

For example, if a linear glissando between 100 and 200 Hz lasts 150 ms, this rule predicts that the pitch perceived is somewhere in the $F_0$ values of the glissando in the time interval 100–150 ms, i.e., between 166.6 and 200 Hz. In Rossi’s terminology, a dynamic tone is a tone with perceived pitch change. Of course, this rule applies only to the experimental situation envisaged in Rossi’s experiments, namely pitch perception measured at the end of tones in vowels. This model was also applied to the pitch of glidellike $F_0$ curves in folk songs. Ross (1987) reported good agreement between his experimental results and Rossi’s rule. In mathematical terms, Rossi’s rule is an example of a time-average (TA) model of pitch perception. Pitch is computed by averaging $F_0$ on the time interval of the tone:

$$p = \frac{\int_{t_1}^{t_2} F_0(\tau) d\tau}{\int_{t_1}^{t_2} d\tau},$$

where $p$ is the pitch perceived, $F_0(\tau)$ is the time-varying fundamental frequency, $t_1$ and $t_2$ are the averaging time limits. In systems for automatic intonation analysis, several authors used “pragmatic” TA models. For instance, Mertens (1987) and House (1990) computed pitch as the average of $F_0$ on 20–30 ms at the end of vowels in syllables.

In a study on pitch perception for short-duration vibrato tones in singing, d’Alessandro and Castellengo (1994) showed that TA models were not suitable in this situation. The main problem was that TA models were not able to give more perceptual weight to the end of tones. They proposed instead a weighted time-average (WTA) model of pitch perception. Pitch is computed as the time average of $F_0$ viewed through a time window:

$$p = \frac{\int_{t_1}^{t_2}(e^{a(\tau-t_2)} + \beta)F_0(\tau)d\tau}{\int_{t_1}^{t_2}(e^{a(\tau-t_2)} + \beta)d\tau},$$

where the constants $\beta$ and $a$ define the raised exponential window. The constant $\alpha$ is the factor of the exponent in the window. It gives more or less weight to the final part of the tone. The constant $\beta$ is the height of the raised exponential window, i.e., the height of a rectangular window, whose width is the tone duration. Contrary to $a$, the constant $\beta$ gives equal weight to all parts of the tone. Thus $\beta$ controls the weight given to the whole tone, while $\alpha$ controls the weight given to the end of the tone. The optimal parameters were $a=2.2$ and $\beta=0.20$. This WTA model was also successfully applied to intonation stylization in speech in d’Alessandro and Mertens (1995).

In summary, the aim of the present study is to address the following matters: (1) to give new measurements of the pitch perceived at the beginning and end of short-duration $F_0$ glissandos; (2) to study the asymmetries in perception of rises and falls; (3) to evaluate a WTA model of $F_0$ integration for pitch perception. This model should be able to predict the pitch perceived at the beginning and end of an $F_0$ glissando. In Sec. I, the experimental method is described. Then the results are presented and analyzed in Sec. II. A numerical model accounting for our data is proposed and discussed in Sec. III. Section IV concludes.

I. METHOD

A. Conditions

The aim of these experiments is to measure the pitch perceived for short-duration $F_0$ glissandos. Linear $F_0$ glissandos are of the form $F_0(\tau)=a\times\tau+b$, where $b$ is the frequency (in Hz) at the beginning of the tone, $a$ is the slope of the tone (if $d$ is the duration of the tone, $a\times d+b$ is the frequency at the end of the tone), and $\tau$ represents time. The stimulus are pairs of synthetic vowels, with a 300-ms silent interval between each member of a stimulus pair. One member of a pair is a glissando (either a rise or a fall). The other member of a pair is flat (constant $F_0$) with the same duration. The test is divided into four parts ABCD, summarized in Fig. 1.
In (A) the first member of a stimulus pair is a rise, and the second member is flat. The subjects are asked to adjust a flat tone to the end of a rise.

In (B) the first member of a stimulus pair is flat, and the second member is a rise. The subjects are asked to adjust a flat tone to the beginning of a rise.

In (C) the first member of a stimulus pair is a fall, and the second member is flat. The subjects are asked to adjust a flat tone to the end of a fall.

In (D) the first member of a stimulus pair is flat, and the second member is a fall. The subjects are asked to adjust a flat tone to the beginning of a fall.

All the stimuli are centered at 220 Hz, a typical F0 for a female voice. Four parameters are studied: the glissando direction (rise/fall), the extremity matched (beginning/end), duration (50, 100, 200, and 300 ms) and extent. The glissando extents are defined according to the glissando threshold, using data reported in ‘t Hart et al. (1990). For the 50 ms duration condition, \( G_r = 0.16/(0.05)^2 = 64 \) semitones/s. The extent of the glissando at threshold \( G_n \) is \( 64 \times 0.05 = 3.2 \) semitones. For this condition, the extents studied in the experiments are 0, 0.8, 1.5, 3, 6 and 12 semitones (or 0, 10.17, 18.74, 38.17, 76.63, and 155.56 Hz). This corresponds approximately to 0.25\( G_n \), 0.5\( G_n \), 2\( G_n \), and 4\( G_n \), in addition to the 0-Hz extent condition. The same values are used for the 100, 200, and 300 ms duration conditions. The extents are symmetric around the center value 220 Hz on a semitone (logarithmic) scale: F0 is within 220 Hz±N/2 semitones, where N is the extent of the stimuli in semitones. As the glissandos are linear functions of time in Hz, the value of 220 Hz is not reached at the center of the stimuli in time because of the wider interval in Hz for higher frequencies. In summary, there are 24 conditions for rises (including four “0 semitone” conditions, which are considered as rises) and 20 conditions for falls. The resulting 88 conditions are summarized in Table I, together with the glissando threshold. Glissandos above and below the glissando threshold are considered in the experiments.

### B. Stimuli

The stimuli were synthetic posterior vowels /a/, computed by a digital parallel formant synthesizer. The center frequencies, bandwidth and relative amplitudes of the first four formants (\( F_1, F_2, F_3, F_4 \)) were typical values for a female speaker: \( F_1 = (650 \text{ Hz, } 80 \text{ Hz, } 0 \text{ dB}), \ F_2 = (1100 \text{ Hz, } 90 \text{ Hz, } -8 \text{ dB}), \ F_3 = (2900 \text{ Hz, } 140 \text{ Hz, } -11 \text{ dB}), \ F_4 = (3300 \text{ Hz, } 130 \text{ Hz, } -20 \text{ dB}) \). Synthetic voice stimuli were chosen because they were fully controlled, and because they were closer to natural sounds (like speech and singing) than pure tones or square waves. The signals corresponding to all the experimental conditions were synthesized and stored on a computer disk. Signals were sampled at 16 kHz, using 16 bits per sample. Prestored examples were played at test time, following the procedure described below.

The signal peak amplitude was the same for all the stimuli. Sound intensity variation within all the stimuli was less than 2.4 dB. This was small enough to avoid consideration of the effect of intensity changes on pitch perception [see Rossi (1978b) for a discussion of interaction of amplitude and fundamental frequency glissandos in pitch perception].

### C. Procedure

Previous data published on the perceived pitch for short-duration F0 glissandos were generally obtained using a forced-choice procedure (Rossi, 1971; ‘t Hart et al., 1990). However, we preferred a method of adjustment similar to the method used in a study of pitch perception for short-duration vibrato tones in d’Alessandro and Castellengo (1994). The method of adjustment has a number of advantages: (1) each response of a subject gives an estimation of the pitch perceived; (2) the method helps concentration, as the subject is actively engaged in the adjustment process; (3) the data obtained are less variable than forced-choice data in frequency discrimination tasks, as shown by Wier et al. (1976).

Subjects were asked to adjust the frequency of the flat tone so that it was equal to the frequency at the beginning or at the end of the glissando. To make an adjustment, the subject pointed to a place on a one-octave scale on the computer screen, pressed a button, and heard a stimulus pair. If she/he was satisfied with this pair, the subject reported a match. If she/he was not satisfied with this pair, the subject moved the mouse pointer on the scale and listened to the corresponding stimulus pair. Low frequencies were on the left of the scale, and high frequencies on the right. For each match, the subject had no time limit for giving an answer, and could play any stimulus pair for the same condition. Subjects typically listened to about ten stimulus pairs while manipulating the frequency of the flat tone before reporting a match. The sub-

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50 ms</td>
<td>100 ms</td>
<td>200 ms</td>
<td>300 ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>E</td>
<td>B</td>
<td>E</td>
<td>B</td>
<td>E</td>
<td>B</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>F</td>
<td>R</td>
<td>F</td>
<td>R</td>
<td>F</td>
<td>R</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>0 ST</td>
<td>b</td>
<td>⋯</td>
<td>b</td>
<td>⋯</td>
<td>b</td>
<td>⋯</td>
<td>b</td>
<td>⋯</td>
</tr>
<tr>
<td>0.8 ST</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
</tr>
<tr>
<td>1.5 ST</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>3 ST</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>6 ST</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>12 ST</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
</tbody>
</table>
TABLE II. Experimental results. Means and standard deviations for the matches reported by the subjects, as a function of extent, duration, rise/fall, and begin/end (see parts ABCD in the text). Means and standard deviations are in Hz.

<table>
<thead>
<tr>
<th>Extent ST (Hz)</th>
<th>50 ms</th>
<th>100 ms</th>
<th>200 ms</th>
<th>300 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Part A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0(220.9)</td>
<td>221.5±1.8</td>
<td>221.1±1.0</td>
<td>221.2±1.0</td>
<td>221.1±1.0</td>
</tr>
<tr>
<td>0.8(215.0–225.1)</td>
<td>224.0±7.1</td>
<td>225.5±6.9</td>
<td>226.2±8.7</td>
<td>227.0±9.5</td>
</tr>
<tr>
<td>1.5(210.6–229.4)</td>
<td>222.8±5.4</td>
<td>227.7±10.6</td>
<td>229.8±9.1</td>
<td>231.4±10.8</td>
</tr>
<tr>
<td>3.0(201.7–239.9)</td>
<td>234.7±15.5</td>
<td>238.6±12.4</td>
<td>241.1±11.6</td>
<td>242.0±12.7</td>
</tr>
<tr>
<td>6.0(185.0–261.6)</td>
<td>255.1±13.7</td>
<td>257.0±9.7</td>
<td>255.2±6.8</td>
<td>256.5±5.3</td>
</tr>
<tr>
<td>12.0(155.6–311.1)</td>
<td>281.9±12.4</td>
<td>283.7±12.8</td>
<td>291.0±13.7</td>
<td>292.0±12.7</td>
</tr>
<tr>
<td></td>
<td>Part B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0(220.9)</td>
<td>218.1±3.6</td>
<td>219.0±1.5</td>
<td>219.1±2.5</td>
<td>219.5±1.4</td>
</tr>
<tr>
<td>0.8(215.0–225.1)</td>
<td>219.0±2.0</td>
<td>218.3±3.4</td>
<td>218.1±2.9</td>
<td>218.1±4.0</td>
</tr>
<tr>
<td>1.5(210.6–229.4)</td>
<td>215.5±9.9</td>
<td>216.3±10.3</td>
<td>216.1±10.7</td>
<td>215.2±10.5</td>
</tr>
<tr>
<td>3.0(201.7–239.9)</td>
<td>213.9±12.4</td>
<td>216.0±13.1</td>
<td>213.3±11.6</td>
<td>212.4±11.3</td>
</tr>
<tr>
<td>6.0(185.0–261.6)</td>
<td>215.7±24.3</td>
<td>213.3±25.4</td>
<td>209.5±20.0</td>
<td>203.8±19.4</td>
</tr>
<tr>
<td>12.0(155.6–311.1)</td>
<td>215.9±39.7</td>
<td>213.8±41.1</td>
<td>208.2±35.7</td>
<td>201.8±29.1</td>
</tr>
<tr>
<td></td>
<td>Part C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8(215.0–225.1)</td>
<td>217.6±7.2</td>
<td>217.1±8.3</td>
<td>217.3±7.2</td>
<td>216.9±4.0</td>
</tr>
<tr>
<td>1.5(210.6–229.4)</td>
<td>217.9±2.0</td>
<td>217.5±2.6</td>
<td>215.3±4.9</td>
<td>214.6±5.4</td>
</tr>
<tr>
<td>3.0(201.7–239.9)</td>
<td>215.2±4.1</td>
<td>213.5±5.1</td>
<td>210.9±6.3</td>
<td>210.4±6.6</td>
</tr>
<tr>
<td>6.0(185.0–261.6)</td>
<td>208.1±15.9</td>
<td>205.7±13.8</td>
<td>201.0±12.2</td>
<td>198.0±10.0</td>
</tr>
<tr>
<td>12.0(155.6–311.1)</td>
<td>214.1±31.2</td>
<td>202.0±22.0</td>
<td>193.7±20.5</td>
<td>193.5±20.5</td>
</tr>
<tr>
<td></td>
<td>Part D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8(215.0–225.1)</td>
<td>220.2±2.0</td>
<td>221.1±2.8</td>
<td>223.1±4.4</td>
<td>222.5±3.0</td>
</tr>
<tr>
<td>1.5(210.6–229.4)</td>
<td>223.7±7.7</td>
<td>224.1±5.3</td>
<td>225.3±7.2</td>
<td>226.7±5.2</td>
</tr>
<tr>
<td>3.0(201.7–239.9)</td>
<td>228.3±7.8</td>
<td>231.1±9.5</td>
<td>234.0±11.4</td>
<td>235.0±6.7</td>
</tr>
<tr>
<td>6.0(185.0–261.6)</td>
<td>246.6±16.5</td>
<td>250.3±14.3</td>
<td>251.6±11.6</td>
<td>254.3±9.5</td>
</tr>
<tr>
<td>12.0(155.6–311.1)</td>
<td>267.4±18.8</td>
<td>272.1±18.9</td>
<td>275.9±22.0</td>
<td>282.0±16.6</td>
</tr>
</tbody>
</table>

II. RESULTS AND DISCUSSION

A. Results

Means and standard deviations averaged for all subjects are reported in Table II and plotted in Fig. 2. For each condition in Fig. 2, horizontal sides of the box represent duration and vertical sides of the box represent extent. The four empty circles near the corners of the boxes represent the means obtained for the four parts ABCD of the test. The coordinates of a circle represent the mean frequency (along the y axis) reported by the subjects, and the time (along the x axis) that corresponds to the instant when the glissando frequency was equal to the mean frequency. Thus the F0 rises can be represented in the boxes by a line between the bottom left corner and the top right corner. The falls can be represented in the boxes by a line between the top left corner and the bottom right corner. If the subjects were able to perfectly perceive the time-varying fundamental frequency, matches would place the four points in the four corners. Part A (rise/end) is generally in the right/top corner, part B (rise/beginning) is generally in the left/bottom corner, part C (fall/end) is generally in the right/bottom corner, and part D (fall/beginning) is generally in the left/top corner. Responses are always close to the extremity matched. This indicates that subjects did not take all the glissando into account, but rather made an average of the frequencies present near the extrem-

D. Subjects

A group of eight musically trained subjects participated in the experiments. One of the subjects was one of the authors (CdA), and the others were music students who were paid for the experiments. Results are likely to be accurate with trained and selected subjects. However, interpretation of the results as representative of the general population may be questionable, because of the superior skills of our subjects in pitch discrimination. With 8 subjects and 4 sessions, 32 matches were made for each of the 88 experimental conditions.
The parts can be classified in the order A→B→C→D, i.e., the higher extremity is the beginning, and farther off the beginning when matched at the beginning. Typically, falls are adjusted closer to the end when matched at the end. Also, standard deviations are smaller for adjustment of the higher extremity (compare A and D, and B and C). For the large extent conditions (6 semitones and 12 semitones), the standard deviations can be ordered as: A<D<C<B. Thus, it seems that when adjustments are further from the extremity matched, they are also less consistent. For the smaller extent conditions, it is difficult to interpret the differences in standard deviation. For the 0 semitone condition, the standard deviations are small when adjusted at the end. In this case, they are close to the difference limens for short tones (Moore, 1973).

An analysis of variance was carried out with factors “duration” (four values), “rise/fall” (two values), “beginning/end” (two values), “extent” (five values), “session” (four values). The factor “duration” was not significant \( F(3,560) = 20.055; p = 0.9830 \). As a matter of fact, only a slight dependence on duration is noticeable in Fig. 2. Means followed a slightly rising or falling line as duration increased for a same extent, but this tendency was not significant. This showed that duration played only a minor role compared to other factors. No significant effect was found for the factor “beginning/end” \( F(1,560) = 1.936; p = 0.1646 \); subject responses averaged for parts A and C were not different from the subject responses averaged for parts B and D. The effect of the factor “rise/fall” was significant \( F(1,560) = 26.94; p < 0.0001 \). This indicates that the asymmetry observed in perception of rises and falls is significant. The factor “extent” was significant \( F(4,560) = 72.67; p < 0.0001 \). This was likely to occur, as for most conditions the extents were close to or above the glissando threshold. Interactions between “rise/fall” and “beginning/end” were very strong \( F(1,560) = 1092.80; p < 0.0001 \). This indicates different behaviors for the different parts A, B, C and D. The interaction between “rise/fall” and “extent” was significant \( F(4,560) = 3.07; p = 0.0162 \). Again, this confirms an asymmetry in perception of rises and falls, which depends on extent. All other interactions were not significant.

**B. Asymmetry between higher and lower extremities**

Rule (1) states that the higher extremity is “better” perceived than the lower extremity. “Better” perceived means that adjustments are closer to the higher extremity, when it is the extremity matched, compared to the lower extremity,
again when it is the extremity matched. This experimental result suggests that the high register might be perceptually more important than the low register. Similar findings have already been reported in other works. A difference between rises and falls was noticed by Rump and Hermes (1996) in an experiment on prominence in intonation. The stimuli were /mamama/ utterances, with fundamental frequency around 100 Hz. They found that high pitch levels contributed to prominence more than low pitch levels. Demany and McAnally (1994) and Demany and Clément (1995, 1997) conducted a number of studies on perception of wide frequency modulations. In these experiments, pure tones or complex tones were used, between 250 and 4000 Hz. They consistently found that frequency peaks are better perceived than frequency troughs. This effect seems robust against factors like the carrier signal, modulation waveform, frequency register and intensity. In a study on perception of contour tones, Hombert (1975) asked subjects to adjust a static tone at the beginning of a rise or a fall. The results are very similar to ours: the beginning of a gliding tone is better perceived for falls than for rises. In other words, for a task rather similar to the present experiments, Hombert (1975) found the same type of effect: The higher extremity is better perceived than the lower extremity. This indicates that rule (1) might be a rather general effect for perception of supraliminar (or wide) frequency modulations.

C. Asymmetry between beginning and end

The second type of asymmetry found in our data, rule (2), is an asymmetry between adjustments at the beginning and at the end of tones. This type of asymmetry has been reported in several studies of wide frequency modulations, as in formant transitions between consonants and vowels in speech. Brady et al. (1961) suggest that a higher perceptual weight could be assigned to the most recent information. More recently, Porter et al. (1991), Wieringen and Pols (1995), Wieringen (1995), and Cullen et al. (1992) reported differences in sensitivity between initial and final transitions of short formant glides. All these studies conclude that the end is better perceived than the beginning. This indicates that rule (2) might be a rather general effect for perception of time-varying frequency modulation.

D. Asymmetries between rises and falls

The possible asymmetry in perception of rises and falls has been the subject of controversial discussion in many papers on tone or intonation perception. Dooley and Moore (1988) discussed asymmetry in difference limens for glide detection. In a two-alternative, forced-choice procedure, they measured frequency difference limens (DLF) for rises and falls. Rises and falls were presented randomly before and after a constant tone signal. Signals were generated, corresponding to parts A and B on the one hand and C and D on the other, and no effects between beginning and end were likely to occur. Pure tones were used, with center frequencies ranging from 500 to 8000 Hz, and durations ranging from 50 to 800 ms. Dooley and Moore (1988) found that UP–DL (DLF for rises) were consistently higher than DOWN–DL (DLF for falls). Thus it seems that rises are more difficult to perceive than falls. The opposite effect is reported by Gardner and Wilson (1979). Dooley and Moore conclude that these differences are difficult to interpret. The differences between rises and falls went unnoticed in many previous works on the glissando threshold, as ’t Hart et al. (1990) pointed out. They also noted the large dispersion of the data on the glissando threshold found in the literature. All these studies were for difference limens. In this case, it is difficult to conclude about possible asymmetries between rises and falls. On the contrary, in the case of wide and supraliminar frequency modulations, it seems that the difference between falls and rises is consistent in many experiments (Hombert, 1975; Rump and Hermes, 1996; Demany and McAnally, 1994; Demany and Clément, 1995, 1997). In this case, the combination of rules (1) and (2) can shed some light on the possible asymmetry in perception of rises and falls, as it is discussed above.

E. Summary

In summary: (1) the mean frequency perceived for short-duration F0 glissandos is in the direction of the extremity matched; (2) distances of adjustments to the extremity differ for rises and falls; (3) adjustments are closer to the extremity for adjustments at the beginning; (4) adjustments are closer to the higher extremity compared to the lower extremity; (5) “duration” plays a minor role; (6) “extent” is highly significant. As a consequence, the four parts ABCD for a given condition of extent and duration have distinct behaviors, which can be seen in the asymmetric butterfly patterns of Fig. 2.

Time averaging of F0 could explain pitch perception for F0 glissandos. It also seems that not all the glissando is taken into account for time averaging. These results are compatible with Rossi’s 2/3 rule. One can easily check that most of our experimental data are comprised between the second and the third third of the glissandos. However, this rule is not really a model, as it gives only an indication of where pitch is. It is of little use in the case of large extent. For example, if one considers the 12 semitone conditions, Ros- si’s rule predicts that the pitch perceived is between the second and the third third of the glissando, and one third of the glissando represents more than 50 Hz. Of course the result is correct, but not very accurate. In the next section, a more accurate model is discussed.

III. NUMERICAL MODEL

A. Equation of the model

A numerical model able to predict experimental data is searched for. The aim of this model is to predict the pitch perceived at the beginning and end of a given F0 glissando. The model should be able to give a response which is close to the subjects’ mean response. Both TA and WTA models are studied. The TA model integrates F0 over a time interval of the glissando. The integration interval is near the extremity matched. Assuming a linear glissando $F0 = a \tau + b$, a TA model is

---

The model has only one parameter, \( \delta \), defined as follows. Let \( d \) be the glissando duration. Then \( \delta \) is the length of the averaging interval for integration, expressed as a proportion of \( d \). The time limits of integration are \((t_1 = 0, t_2 = \delta \times d)\) for an adjustment at the beginning, and \((t_1 = d - \delta \times d, t_2 = d)\) for an adjustment at the end. The TA model is computed as

\[
P(d) = \frac{\exp(a(t-1)) \exp(\alpha t) \exp(\alpha t) d \tau}{\int_1^2 \exp(a(t-1)) \exp(\alpha t) d \tau} = \frac{a \tau - a \tau}{1 - \exp(-a \delta \tau)} + b - \frac{a}{\alpha}.
\]

Let \( d \) be the glissando duration. Then the time limits of integration are \((t_1 = 0, t_2 = \delta \times d)\) for an adjustment at the beginning, and \((t_1 = d - \delta \times d, t_2 = d)\) for an adjustment at the end. The WTA model is computed as:

\[
P(\alpha, \delta) = \frac{\int_1^2 \exp(\alpha(t-1)) \exp(\alpha t) d \tau}{\int_1^2 \exp(\alpha(t-1)) \exp(\alpha t) d \tau} = \frac{a \delta + b}{1 - \exp(-a \delta \tau)} + b - \frac{a}{\alpha}.
\]

The quality \( Q(\delta, \alpha) \) of the model with parameters \( \delta, \alpha \) is the number of conditions for which the model and the data are statistically equivalent:

\[
Q(\delta, \alpha) = \sum_{n=1}^{N} q_{\delta, \alpha}(n),
\]

where \( N \) is between 1 and \( N \), and where \( N \) is the number of conditions considered.

The minimum rms distance between the model and data, using an iterative minimization procedure, gives the optimal parameters. However, the raw distance between the model and the data is not very informative by itself. A good way to gain further insight into the model quality is to evaluate the fit to the data, by computing the number \( Q \) of conditions for which the model response is not statistically different from the subject’s response: The larger the \( Q \), the better the fit. \( Q \) is computed as follows. Let \( e(n) \) be the experimental standard deviation (reported in Table II) for a particular condition \( n \) of factors “duration,” “extent,” “rise/fall,” and “beginning/end.” The statistical equivalence between this predicted value and the experimental data is computed by the Student \( t \)-test (recall that 32 responses are available for each test condition): \( t(n) = [(m(n) - p_{\delta, \alpha}(n)) / \sqrt{31}] / e(n) \). Using a level of significance of 0.05, the critical value is 2.03 for a two-tailed test with \( df = 31 \). The statistical equivalence between the predicted value and the experimental data for a condition \( n \) is computed using a function \( q_{\delta, \alpha}(n) \) defined by

\[
q_{\delta, \alpha}(n) = \begin{cases} 
1, & \text{if } -2.03 < t(n) < 2.03, \\
0, & \text{otherwise.}
\end{cases}
\]

The quality \( Q(\delta, \alpha) \) of the model with parameters \( \delta, \alpha \) is the number of conditions for which the model and the data are statistically equivalent:

\[
Q(\delta, \alpha) = \sum_{n=1}^{N} q_{\delta, \alpha}(n),
\]
C. Discussion

The TA and WTA models are able to predict 85% and 94% of the data, provided that the four parts ABCD are considered separately. The percentages are reduced to only 61% and 65% with the same parameter setting for all parts. This result is not surprising, given the asymmetries observed in the experimental data. Thus for modeling the four parts ABCD must be considered separately. In this case, good results are obtained for both models, though the WTA model performs better than the TA model.

The optimal $\alpha$ and $\delta$ parameters for the WTA model are different for the four parts of the test. The interpretation of $\delta$ is straightforward: It represents the proportion of the glissando over which integration occurs. Interpretation of $\alpha$ is two-fold. On the one hand, the sign of $\alpha$ indicates if more attention was paid to the beginning or end of the glissando. A positive $\alpha$ indicates that a higher perceptual weight is assigned to the most recent information (end). A negative $\alpha$ indicates that a higher perceptual weight is assigned to the most distant information (beginning). On the other hand, the magnitude of $\alpha$ represents the amount of damping introduced in averaging. A large magnitude means that a higher perceptual weight is assigned to the information close to the extremity matched. A small magnitude means that all the parts of the time interval used for averaging have almost the same weight. The averaging interval ($\delta=0.26$) is very short for part A: about 26% of the end of the glissando. The parameter $\alpha=24.7$ is positive, which means that the end of the glissando has more weight. The averaging window length in part B ($\delta=0.82$) is long: about 82% of the glide. The parameter $\alpha=-7.5$ is negative. This indicates that more attention was paid to the beginning of the glissando. The same remark holds for part D, where $\alpha=-19.4$ is negative, though for part D the averaging window ($\delta=0.56$) is shorter than for part B. For part C, the parameter $\alpha=13.5$ is positive, which indicates that more attention was paid to the end of the glissando. The averaging window ($\delta=0.75$) is long, about 75% of the tone duration. The same remarks hold for the TA model. The averaging window is shorter for part A (18%), longer for part B (69%). It is also long for parts C and D.

The combination of the magnitude of $\alpha$ and $\delta$ can be used for computing the distance between the model responses and the extremity matched. Recall that in the results section, it was found that this distance is ordered as $A<D<C<B$ for the experimental means. The constant $\alpha$ magnitude represents the damping factor. When it is high, damping is high and then the model response is close to the extremity. The distance of the model response and the extremity can be ordered as follows: $A<D<C<B$ because the corresponding $\alpha$ are $24.7(A)>19.4(D)>13.5(C)>7.5(B)$. The parameter $\delta$ represents the proportion of the glissando used for integration. When $\delta$ is small, the model response is close to the extremity. We obtain the same order: $A<D<C<B$, because $0.26(A)<0.56(D)<0.75(C)<0.82(B)$. Therefore, the optimal parameters obtained for the WTA model are in good agreement with the heuristic rules derived

<table>
<thead>
<tr>
<th></th>
<th>TA $\delta$</th>
<th>$Q$ (#)</th>
<th>%</th>
<th>WTA $\delta$</th>
<th>WTA $\alpha$</th>
<th>$Q$ (#)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>All data</td>
<td>0.48</td>
<td>49 (80)</td>
<td>61</td>
<td>0.54</td>
<td>7.2</td>
<td>52 (80)</td>
<td>65</td>
</tr>
<tr>
<td>Part A</td>
<td>0.18</td>
<td>17 (20)</td>
<td>85</td>
<td>0.26</td>
<td>24.7</td>
<td>17 (20)</td>
<td>85</td>
</tr>
<tr>
<td>Part B</td>
<td>0.69</td>
<td>20 (20)</td>
<td>100</td>
<td>0.82</td>
<td>-7.5</td>
<td>20 (20)</td>
<td>100</td>
</tr>
<tr>
<td>Part C</td>
<td>0.59</td>
<td>16 (20)</td>
<td>80</td>
<td>0.75</td>
<td>13.5</td>
<td>20 (20)</td>
<td>100</td>
</tr>
<tr>
<td>Part D</td>
<td>0.45</td>
<td>15 (20)</td>
<td>75</td>
<td>0.56</td>
<td>-19.4</td>
<td>18 (20)</td>
<td>90</td>
</tr>
<tr>
<td>ABCD</td>
<td>...</td>
<td>68 (80)</td>
<td>85</td>
<td>...</td>
<td>...</td>
<td>75 (80)</td>
<td>94</td>
</tr>
</tbody>
</table>
from the experimental results: The same order is found for the four parts ABCD of the test.

In summary, it seems that both the TA and the WTA models are able to predict a significant amount of the experimental data. However, the WTA model performs better than the TA model. This is not surprising, since the WTA model has two parameters compared to only one parameter for the TA model. The model’s parameters give some insight into the experimental data, and are compatible with the butterfly patterns found in the experimental data. It must be pointed out that it is necessary to know the glissando direction and whether the match is at the beginning or end in order to select a set of parameters. This means that the WTA model in itself is not able to explain the asymmetries, but must be adapted to each situation. In this case it is able to predict the pitch perceived. The WTA model is a functional model, and not a model of the physiological or psychological processes that take place in pitch perception. It may help to compute the pitch perceived for an F0 glissando, but is not able to explain how this pitch is perceived. In particular, the explanation of experimental asymmetries in pitch perception is still unclear.

IV. CONCLUSION

In this study, we studied perception of short-duration linear F0 glissandos. Perception of rising and falling glissandos, matched at the beginning and at the end, was measured. Significant differences between the four main parts ABCD of the experiments (rises adjusted at the end, rises adjusted at the beginning, falls adjusted at the end, falls adjusted at the beginning) were found. This indicates that there is a difference between perception of rises and falls and perception of the beginning and the end of tones. This might be explained by a greater perceptual weight for the high register, compared to the low register. The experimental results showed that some sort of averaging of the F0 contours was performed by the subjects. Frequencies near the extremity matched (beginning or end) seemed more important. Simple numerical models for prediction of the experimental data were tried: a time-average model and a weighted time-average model of pitch perception. Optimal parameters for these models were estimated using a statistical distance criterion. The models performed well in predicting experimental data: 94% of the data computed with the best WTA models were not significantly different from experimental data. The optimal model’s parameter settings varied significantly among parts ABCD (rises versus falls, adjustment at the end versus adjustment at the beginning). This was consistent with the finding of significant differences in the perception of the four parts. It can be concluded from our experiments that pitch perception for short-duration F0 glissandos corresponds to a weighted time average of F0. The WTA model proposed in the present study is close to the WTA model that was proposed for perception of short-duration vibrato tones in d’Alessandro and Castellengo (1994). However, the parameters obtained in both studies are different, as in this study they are also different among different parts.

ACKNOWLEDGMENTS

We are grateful to John P. Madden and Adrian Houtsma for their detailed constructive comments on an earlier version of this article, and to the editorial assistance and criticisms provided by D. Wesley Grantham. Thanks also to Olivier Piot, who helped in the experimental data collection.

1Mario Rossi, not to be confused with Jean-Pierre Rossi, one of the authors.
2Rossi’s rule was stated for adjustment at the end of stimuli. The same type of rule can be formulated in the case of adjustment at the beginning, as follows: For dynamic tones in a vowel, the pitch perceived at the beginning corresponds to a point between the first and the second third of the vowel.