Transitional Characteristics of Fundamental Frequency in Singing

Hiroki Mori†, Wakana Odagiri†, Hideki Kasuya†, and Kiyoshi Honda‡

†Faculty of Engineering, Utsunomiya University
‡ATR Human Information Science Laboratories

Abstract
The $F_0$ characteristics during the note transition period are investigated. Analyses of the singing voice of a professional singer strongly suggest that apparent asymmetries exist in the mechanisms used for controlling rising and falling. Specifically, the $F_0$ contour in rising transitions can be modeled as the step response from a critically-damped second-order linear system, whereas that in falling transitions can be modeled as the step response from an underdamped second-order linear system with fixed transition time.

1. Introduction
When a singer is about to move to the next note, a rapid $F_0$ change is thought to occur, which is a consequence of an $F_0$ control system whose target value is changing. This study set out to establish a mathematical model to describe $F_0$ dynamics in singing. To accomplish this, transitional characteristics of $F_0$ were investigated in detail based on the recordings of passages sung by a professional vocalist in Western classical style.

2. Analysis
2.1. Experimental Procedure
A professional baritone (age 49) was presented with 24 independent musical scores, which were the combination of 2 patterns and 12 intervals. The patterns consisted of “Rise→Fall (D3#)” and “Fall→Rise (C4),” in which the keynote was D3# (155.6 Hz) and C4 (261.6 Hz), respectively. Each score consisted of three notes, where the second note was surrounded by one of the keynotes. Therefore, each score contained one rising and one falling. The interval between the keynote and the second note ranged from 1 to 12 semitones (100–1200 cents). In addition, an extra set of “Fall→Rise (E3)” was presented with keynote E3 (164.8 Hz) and intervals ranging from 1 to 5 semitones. The singer sang each of the 29 scores three times. The singing style was legato, that is, transitions were smooth and natural.

2.2. Parameters and Measurement
From the extracted $F_0$ contours, the following parameters were measured:

steady-state average The averaged value of $F_0$ over the stationary part of each note. The stationary part was identified by visually checking the $F_0$ contour.

interval The difference of the steady-state average between consecutive notes. It is not necessarily the same as the interval prescribed in the score.

transition time The duration of transition. The endpoints of a transition were defined as local extrema of $F_0$ just before (or after) the transition.

overshoot extent The $F_0$ difference between the transition endpoint just after the transition and the steady-state average after the transition.

It is not easy to identify the transition endpoint because the direction of $F_0$ change demanded by the musical score and the one to which the vibrato is going to move may conflict. In such conflicting cases, vibrato can “ride” on the rapid $F_0$ change rather than being interrupted. Transition endpoints in the riding cases were marked according to a predetermined rule.

2.3. Results
Figure 1 shows the measured transition time for each transition. A positive value of interval corresponds to rising, and a negative value to falling. The result reveals that the transition time in the falling cases was virtually constant, whereas in the rising cases, the transition time had a strong positive correlation with interval. There was no difference between the distribution of Rise→Fall pattern and Fall→Rise pattern. Also, no difference was observed in transition time between the Fall→Rise (C4) group and Fall→Rise (E3) group.

For most cases, the transition time for falling was not longer than that for rising. This overall tendency does not conflict with prior studies by Sundberg[1] or by Fujisaki et al.[2].

Figure 2 shows the extent of overshoot measured for each transition. In most cases for rising, the overshoot extent was quite small (< 1 semitone), whereas there was a strong positive correlation between intervals and overshoot extents in the falling cases. A slight difference was observed in the distributions among the patterns (Rise→Fall (D3#), Fall→Rise (C4) and Fall→Rise (E3)), but the overall tendencies were the same.

Previous studies[3, 4] showed that overshoot extent is larger in falling transitions than in rising transitions.
Our data shown in Fig. 2 strongly suggest that different mechanisms are at work in global $F_0$ control for different transition directions.

In the falling cases, most (absolute) overshoot extent was significantly higher than the vibrato depth. Thus, overshoot is almost certainly a separate phenomenon from vibrato in falling transitions. In rising transitions, on the other hand, the overshoot extent is quite small.

Based on the analyses above, it can be concluded that the $F_0$ contour in rising transitions can be modeled as a step response from a critically-damped second-order linear system, whereas that in falling transition can be modeled as a step response from an underdamped second-order linear system with fixed transition time.

Besides $F_0$, there are several acoustic parameters (e.g. voice source amplitude) that dominate the quality of synthesized singing voice. Transitional aspects of such parameters should be studied in future investigations.

### 5. References


