TIMING AND ACCURACY OF FUNDAMENTAL FREQUENCY CHANGES IN SINGING

Guus de Krom and Gerrit Bloothooft
Research Institute for Language and Speech, University of Utrecht, the Netherlands

ABSTRACT

This study deals with relations between the musical score and the acoustic F_0 pattern measured in a sung passage. Four trained singers performed a song in which magnitudes and directions of pitch changes varied systematically. The songs were recorded on the vowels /i/, /a/, and /u/, at three different tempi. Timing differences in the F_0 transitions and deviations from the target values were investigated with a Dynamic Time Warping procedure.

INTRODUCTION

For a singer, the musical score indicates how and when vocal pitch should be varied. However, there exists no one-to-one relation between the prescribed (discrete) note sequence and the F_0 pattern measured in a recorded song, due to for instance F_0 vibrato and inertia of the organic structures involved in phonation. Also, singers have an expressive freedom, which allows them to deviate to some extent from the norm. It may be expected that discrepancies between the pattern of note sequences prescribed in the musical score and the actual F_0 patterns measured in a sung passage depend on the rate at which note sequences are sung (larger deviances at fast tempi), as well as the magnitude of the F_0 difference between successive notes (larger deviances on large intervals).

METHODS

Material

The material used in this study consisted of a song (composed by G. Bloothooft), in which magnitudes and directions of note transitions were systematically varied. The transitions of interest always followed on a particular sequence of "leader" notes (F3-A3 [170-220 Hz] for males, and F4-A4 [340-440 Hz] for females). We chose a fixed leader sequence in order to minimise variations in the immediately preceding context. Nine different transitions were distinguished on the basis of the interval and direction of the steps (see Table 1).

Table 1. Interval magnitude (in semitones) and direction (- downward, + upward) for nine types of transitions.

	1	2	3	4	5	6	7	8	9
interval	-7	-5	-4	-2	+1	+3	+5	+7	+8

Each singer produced nine (legato) versions of the song, using the Dutch vowels /i/, /a/, and /u/, and a slow, medium, and fast tempo (2, 4, and 6 notes per second, respectively).

Recordings and acoustic analyses

The recordings were made in a large, sound treated room. The singers (two males, two females) were standing upright, with the musical score mounted on a stand in front of them. The singers had synthesised piano accompaniment presented over headphones at a comfortable loudness level (simple chords at each measure). A condenser microphone was placed at about 50 cm from the mouth of the singer. The microphone signal was recorded on a DAT recorder.

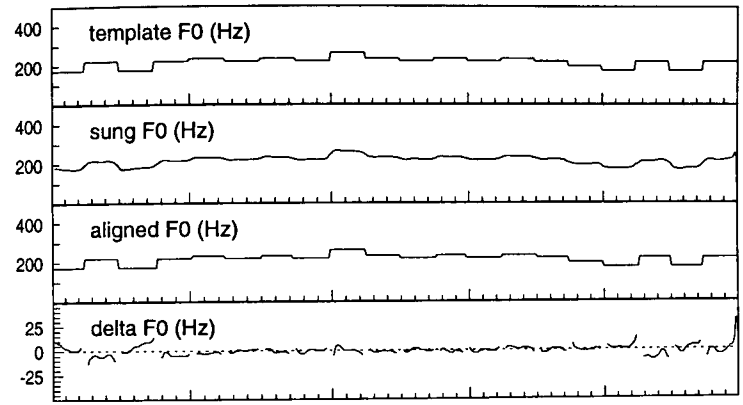

Figure 1. Example of DTW output traces (male singer, medium tempo, duration 5.0 s). From top to bottom, the traces represent the musical score (= template F_0), sung F_0, a trace with the optimal alignment of template F_0 to sung F_0 (aligned F_0), and a trace with the difference between sung F_0 and aligned F_0 (delta F_0).

The DAT recorded songs were transferred to a computer and downsampled to 20 kHz. F_0 was measured with a time-domain algorithm [1], yielding one value (in Hz) for every 10 ms. The musical score was translated into a similar time-by-F_0 format, yielding six template files (3 tempi × 2 versions [for males and females]).

Dynamic Time Warping (DTW)

Relations between the template F_0 traces prescribed in the musical score and sung F_0 traces were investigated by means of a Dynamic Time Warping (DTW) procedure [2]. DTW analyses were performed on the entire lengths of the traces, yielding template F_0, sung F_0, and a DTW-aligned F_0 trace as the optimal match of template F_0 to sung F_0 (see Figure 1).

Parameter definitions

The DTW output files were further processed to obtain measures describing the relative timing of the note transitions and the accuracy of the transitions.

Deviations in the ***timing*** of transitions were investigated by comparing transition instants in the aligned F_0 and template F_0 traces. Three parameters were examined: (1) the transition **lag,** defined as the time difference between the transition instants in the template F_0 and aligned F_0 traces. Negative **lag** values indicate that a transition was made at a later instant than prescribed in the musical score. (2) The absolute value of lag (|**lag**|) yielded another timing parameter. (3) The transition **duration** was defined as the time difference between local maximum or minimum values in the sung F_0 trace just before and after the transition moment.

Parameters indicative of the ***accuracy*** of the transitions were determined by comparing the sung F_0 and aligned F_0 traces. Because these have an identical time basis, a subtraction of sung F_0 from aligned F_0 results in a trace with the instantaneous F_0 deviation in Hz (delta F_0). The local minimum (falling transitions) or maximum (rising transitions) in the sung F_0 trace shortly following the transition moment in the aligned F_0 trace was sought. The difference between this local minimum or maximum and the prescribed value was calculated and expressed in semitones (**F_0dev**, positive values indicating that the singers' F_0 was too high). A second accuracy parameter was defined as the absolute value of F_0dev (**$|F_0\mathrm{dev}|$**).

RESULTS

Analyses-of-variance were performed with **lag, $|$lag$|$, duration, F_0dev, and $|F_0\mathrm{dev}|$** as dependent variables, and interval, tempo, vowel, and the singer's sex as factors (SPSS procedure MANOVA [3]). Because the levels of the interval factor were spaced unequally (see Table 1), polynomial contrasts were applied to these levels. Main effects and two-way interactions were investigated.

Timing parameters

As could be expected, tempo had a significant effect on **lag** ($p < .001$), with values of -276, -80, and -64 ms for the slow, medium, and fast tempo, respectively. Sex ($p = .008$), tempo ($p < .001$), and interval ($p = .039$) had a significant effect on **$|$lag$|$**, as did the sex × tempo interaction ($p = .025$). Figure 2 gives the effect of size on **$|$lag$|$**.

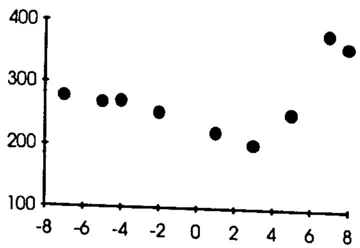

*Figure 2. Mean **$|$lag$|$** in ms as a function of the transition interval in semitones.*

As Figure 2 shows, **$|$lag$|$** decreased from an interval of -7 semitones to an interval of +3 semitones. **$|$lag$|$** was longest for intervals of +7 and +8 semitones, respectively. Thus, it seems that note steps that involve large (especially upward) F0 transitions give rise to large deviations in the timing of the note transitions.

For both males and females, mean **$|$lag$|$** decreased with increasing tempo. For females, the largest difference was found between the slow tempo on the one hand, and the medium and fast tempi on the other (384 ms, versus 176 and 154 ms). For males, these values were 370, 313, and 208 ms, respectively.

Sex, tempo, and interval had significant effects on the **duration** of the transitions ($p < 0.001$, all factors), with significant interactions for sex × vowel ($p = .010$), sex × tempo ($p = .015$), and interval × tempo ($p = .001$). Figure 3 gives the interval × tempo data.

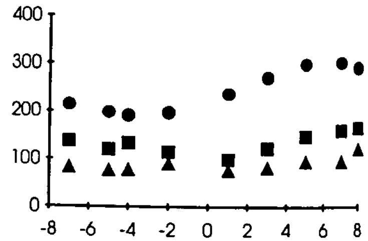

*Figure 3. Mean transition **duration** in ms as a function of the transition interval in semitones. Data are given for slow (circles), medium (squares) and fast tempi (triangles).*

Variations in **duration** related to interval size were most outspoken for the slow tempo. **Duration** was not much influenced by interval size for the medium and fast tempi.

Overall, the **duration** of transitions was shorter for females than for males. The difference was some 30 ms in the slow and medium tempi. No difference was found in the fast tempo.

For the vowels /a/ and /u/, females had a shorter mean transition **duration** than males (135 versus 168, and 130 versus 169 ms). For the vowel /i/, female mean transition duration was slightly longer (154 versus 149 ms).

Accuracy parameters

Interval had a significant effect on mean **F_0dev** ($p < 0.001$). As can be observed in Figure 4, the F0 of the note following the transition was on average too low for downward intervals. A less systematic pattern was found for upward steps, although F_0 was on average slightly too high for intervals between +1 and +5 semitones. These data suggest that the singers exaggerated the prescribed pitch transitions.

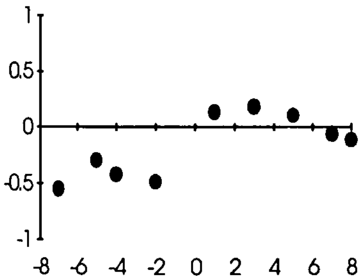

*Figure 4. Mean **F_0dev** in semitones as a function of the transition interval in semitones.*

Interval also had a significant effect on **$|F_0\mathrm{dev}|$** ($p < 0.001$), with a significant (but not systematic) interval × sex interaction ($p < 0.001$). Figure 5 gives **$|F_0\mathrm{dev}|$** data for different intervals.

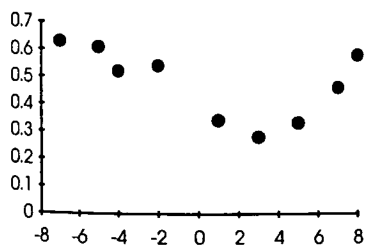

*Figure 5. Mean **$|F_0\mathrm{dev}|$** in semitones as a function of the transition interval in semitones.*

Figure 5 shows that the largest (absolute) deviations in F_0 were found for the largest downward or upward intervals, indicating that these were more difficult to produce than sequences with minor pitch changes. If we compare the **F_0dev** and **$|F_0\mathrm{dev}|$** data, it appears that F_0 of notes following downward transitions was on average too low (negative **F_0dev** values), while F_0 of notes following upward transitions deviated in a more random fashion from the target value (**F_0dev** approximately zero, **$|F_0\mathrm{dev}|$** nonzero).

CONCLUSIONS

We found that the (absolute) interval of note transitions had an influence on the timing of note transitions, as well as the accuracy of the actual F_0 values.

Transition timing was most variable for note sequences that involved large pitch transitions. In the slow tempo, the duration of transitions was longer for upward transitions than for downward transitions.

Systematic F_0 overshoot occurred with downward intervals (more overshoot for larger intervals). Upward intervals resulted in more random F_0 deviations. Tempo and vowel type had no effect on the F_0 accuracy measures.

All four singers reported having difficulties in singing at the fast tempo. Tempo had an effect on the timing data, but not on the accuracy measures. We might therefore tentatively conclude that singers tried to compensate the difficulties encountered in the singing of fast note sequences by adjusting their timing of these transitions.

REFERENCES

[1] Reetz, H. (1989). A fast expert program for pitch extraction. *Proceedings Eurospeech '89*, 476-479.
[2] Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimalization for spoken word recognition. *IEEE ASSP, 26*, 43-49.
[3] Norušis, M.J., (1985). *SPSS-X Advanced Statistics Guide*. McGraw-Hill, New York