

Estimating the Time to Reach a Target Frequency in Singing

Sean Hutchins^a and David Campbell^b

^a*Department of Psychology, McGill University, Montreal, Quebec, Canada*

^b*Department of Statistics and Actuarial Science, Simon Fraser University, Surrey, British Columbia, Canada*

The ability to match pitches quickly and accurately is essential for proficient singing. We describe a new technique for estimating the time to reach a target frequency that uses adaptive optimal-kernel (AOK) time-frequency representations, designed to optimize the time-frequency tradeoff at each time point. We show in two experiments that this measure is more sensitive to tonal priming effects than an onset latency measurement. This analysis is applied to the vocal productions of untrained singers to reveal effects of tonality and pitch height.

Key words: singing; time-frequency representation; priming

Introduction

The ability to match pitch is essential for proficient singing. Pitch-matching ability is one of the most important determinants of singing ability. Whereas many people claim to be “tone deaf,”¹ recent studies show that many people may underestimate their musical proficiency, at least when it comes to pitch matching.^{2,3} These studies have found a wide range of individual difference in proficiency among occasional singers. Natural singing abilities span the range from professional singers, who are consistently more accurate at pitch-matching tasks than untrained singers,⁴ to congenital amusics.⁵ This wide range of abilities contrasts with the other major form of vocal communication, speech, in which all neurologically normal adults without major hearing deficiencies develop high-functioning proficiency.

Hutchins and Palmer⁶ used pitch matching to show priming advantage for the tonic tone (the most important and most frequent tone

in a melody), as well as a repetition priming advantage for individual pitches. This study used a singing task: musically untrained participants heard a short melody, and sang back the pitch of the final melodic tone as soon as possible after hearing it on the syllable /ba/. All participants were female, to control for vocal range. The onset latency (OL) of the sung tones was measured, defined as the time from the onset of the final melodic stimulus tone to the onset of the first pitched part of the singer’s response. Hutchins and Palmer⁶ found a significant effect of tonal priming, such that the onset of responses to melodies were on average 22 ms faster for those that ended on the highly expected tonic tone than those that ended on a less-expected nontonic tone.

We present an alternative metric to the OL measurement, the time to reach the target frequency (TRTF). This measure considers not only the onset of the tone but also the onset of the target frequency, which may occur at or after the OL. The TRTF incorporates both latency and accuracy in measurements of tones whose pitch can change across time. We hypothesize that there should be a tonal priming effect in the TRTF as well.

Address for correspondence: Sean Hutchins, Department of Psychology, McGill University, 1205 Dr. Penfield Drive, Montreal, Quebec, H3A 1B1, Canada. Sean.Michael.Hutchins@umontreal.ca

Determining the TRTF requires both time and frequency resolution, which presents a challenge. To the degree that a frequency is precisely specified, its resolution in time becomes blurred, and likewise as units of time are shortened, one loses accuracy in determining the frequency. The rate at which sung tones move through pitch space affects the ideal ratio between frequency and temporal resolution. To alleviate this problem, we use an adaptive windowing time-frequency estimation method, adaptive optimal-kernel (AOK) time-frequency representations (T-FR),⁷ which adjusts the tradeoff between time and frequency to account for how quickly the signal is changing at any given time. In addition, this method allows us to examine the way singers correct an inaccurate initial pitch, and their trajectory through pitch space.

Onset and TRTF

The AOK T-FR was performed using MATLAB and software available from the site <http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=11551>, using $\kappa = 2400$ (down-sampling the recording to 4900 Hz), $\delta = 128$, and $\alpha = 2$. The ridge was defined as the largest amplitude frequency feature from the AOK T-FR at each point in time after restricting to frequencies less than 500 Hz.

The OL measurement was defined as the earliest time in a vocal response where the following criteria were met:

- The ridge lies within the female register of approximately 175–480 Hz.
- The ridge persists for at least 50 ms within the female register.
- The pitch exceeds a minimum amplitude equal to 10% of the maximum amplitude of the response.

The TRTF measurement was defined as the earliest time in a vocal response where:

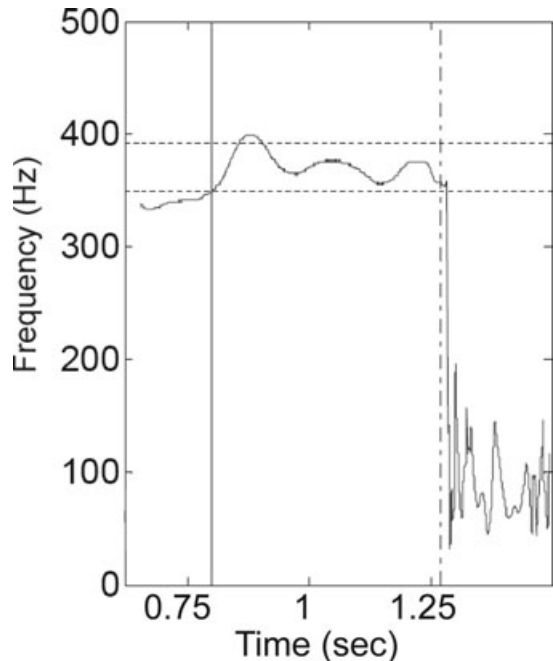


Figure 1. An example of the ridge obtained from the AOK T-FR of a sample vocal production from Hutchins and Palmer.⁶ The vertical line indicates the measured time to target frequency, the dotted-dashed line indicates the signal end, where the signal amplitude drops below 10% of the maximum signal amplitude, and the dashed horizontal lines indicate the pitch tolerance boundaries (one semitone above and below the target frequency).

- The ridge is within one semitone of the target frequency.
- At least 80% of the remaining sung response is within one semitone of the target frequency.
- The OL criteria are met.

Figure 1 shows an example of the ridge obtained from a sample vocal production from Hutchins and Palmer.⁶ The solid vertical line indicates the measured time to target frequency, the dotted-dashed vertical line indicates the signal end, where the signal amplitude drops below 10% of the maximum signal amplitude, and the dashed horizontal lines indicate the pitch tolerance boundaries (one semitone above and below the target frequency).

Results

Comparing OL with TRTF

A 2×2 repeated-measures ANOVA comparing the factors of measurement type and tonic ending showed a main effect of measurement type, $F(1, 23) = 50.85$, $P < 0.001$. The TRTF measurement yielded significantly longer latencies than the OL measurement ($OL \leq TRTF$ by definition). There was also a main effect of tonic ending, $F(1, 23) = 19.97$, $P < 0.001$, as well as an interaction between measurement type and tonic ending, $F(1, 23) = 57.78$, $P < 0.01$. Participants responded faster to tonic endings than nontonic endings across both measurement types, and this effect was larger for the TRTF measurement (33 ms) than for the OL measurement (22 ms).

To ensure that this pattern of results was not specific to this particular experiment, the same analyses were also applied to Hutchins and Palmer's⁶ Experiment 3, which used an identical task and design (only the timbre of the final target tone was changed) with a different group of singers. The 2×2 repeated-measures ANOVA showed the same pattern of results as Experiment 2, with main effects of measurement type, $F(1, 23) = 57.78$, $P < 0.001$, tonic ending, $F(1, 23) = 13.43$, $P < 0.01$, and an interaction between measurement type and tonic ending, $F(1, 23) = 4.98$, $P = 0.04$.

Comparing the Paths Taken to Reach the Target Frequencies

The ridges obtained from the AOK T-FR give a description of the trajectories of the pitches (fundamental frequency) of the sung responses. Figure 2 shows the range of trajectories in the 400 ms after the OL for the data from Experiment 2 for each target pitch (C4-G4) across all responses. The dark lines indicate the median ridge trajectory, and the lines outside of those indicate the 80%, 90%, and 95% trajectories. The median trajectories for all tar-

gets began near the middle of the vocal range, and fell to reach low targets, but ascended to reach higher target pitches. A repeated-measures ANOVA comparing the factors of measurement type and target pitch revealed a main effect of target pitch, $F(7,17) = 8.68$, $P < 0.001$, and an interaction between target pitch and measurement type, $F(7,17) = 4.43$, $P < 0.01$. Both response latencies and initial inaccuracy times (differences between the TRTF and OL measurements) were longer on average for high and low target pitches compared with mid-range target pitches.

Discussion

The AOK T-FR measurement of the time for singers to reach target frequency showed a larger effect of tonal priming than could be obtained with traditional OL measurements.⁶ The TRTF measure yielded tonal priming effects (measured as the difference in response latencies between tonic and non-tonic conditions) that were approximately 50% larger than the OL measures. Singers were less accurate in initial frequency and took longer to adjust to the correct pitch when the pitch they were attempting to match was a nontonic tone than when it is a tonic tone. Thus, the TRTF measure shows that tonic priming is not limited solely to planning of the vocal production, but continues to have an effect even after the tone has begun to be produced.

The ridges provided by the AOK T-FR also provide a useful estimate of the average vocal trajectory that singers use to reach a target pitch. Singers tended to start in the mid-range of frequencies regardless of the pitch of the target tone, and adjusted their frequency upwards to reach high-frequency target pitches, and downward to reach low-frequency target pitches. This provides further evidence that, in speeded pitch-matching paradigms, such as those of Hutchins and Palmer,⁶ participants continue to tune their vocal responses even after beginning to sing. One interesting

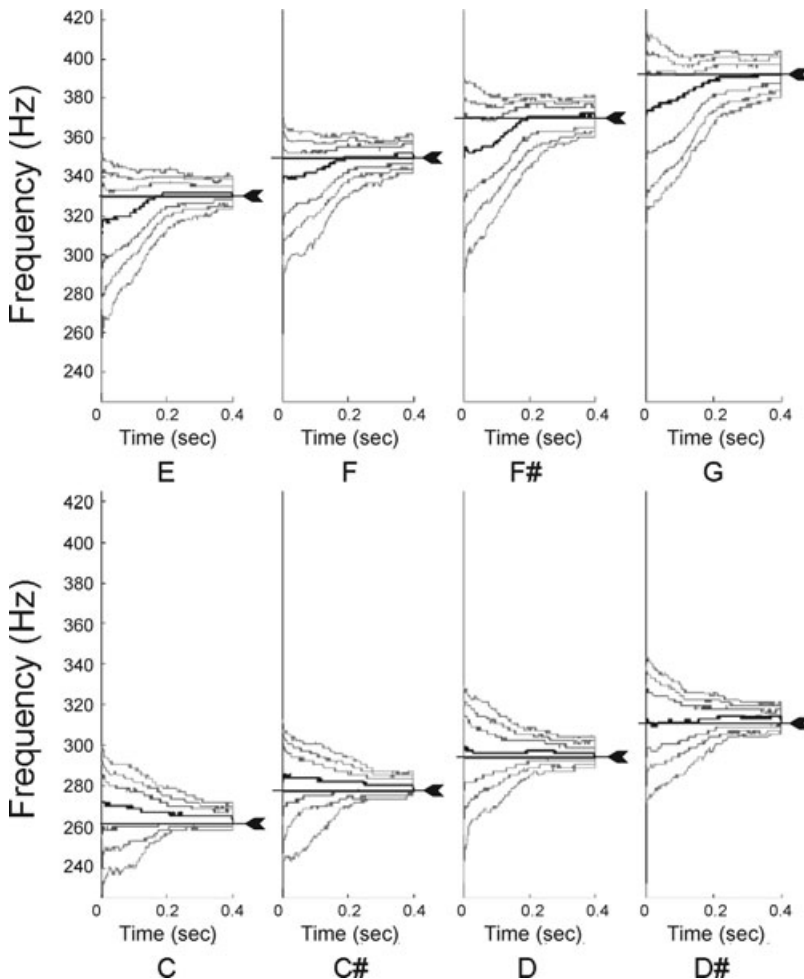


Figure 2. The range of ridge trajectories in the 400 ms after the OL for the data from Experiment 2 for each target pitch (C4-G4) across all responses. The dark lines indicate the median ridge trajectory, and the lines outside of those indicate the 80%, 90%, and 95% quantile interval boundaries. The horizontal lines with arrows indicate the target pitch (referred to by musical pitch below each graph).

feature of these ridges is that they begin to move toward the target frequency immediately, indicating that participants are not waiting to receive their own external feedback. This may indicate that the onset of singing occurs while the vocal folds are still in the process of adjusting.

In sum, the AOK T-FR provides a useful way of analyzing sung vocal production. It can yield stronger measurements of psychological processes, such as tonal priming, than can be found solely with OLs and gives a more com-

plete picture of how singers change pitch over the course of time.

Acknowledgments

We would like to thank Caroline Palmer, Jim Ramsay, Janeen Loehr, and Werner Goebel. This research was supported in part by a Tomlinson Fellowship and a Le Fonds Québécois de la Recherche sur la Nature et les Technologies Fellowship to S.H. D.C. was supported by

MITACS Grant 208683 and NSERC Grant 224607.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Sloboda, J.A., J.A. Wise & I. Peretz. 2005. Quantifying tone deafness in the general population. *Ann. N. Y. Acad. Sci.* **1060**: 255–261.
2. Dalla Bella, S., J.-F. Giguère & I. Peretz. 2007. Singing proficiency in the general population. *J. Acoust. Soc. Am.* **121**: 1182–1189.
3. Pfordresher, P.Q. & S. Brown. 2007. Poor-pitch singing in the absence of “tone deafness.” *Music. Percept.* **25**: 95–115.
4. Murry, T. 1990. Pitch-matching accuracy in singers and nonsingers. *J. Voice* **4**: 317–321.
5. Ayotte, J., I. Peretz & K. Hyde. 2002. Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* **125**: 238–251.
6. Hutchins, S. & C. Palmer. 2008. Repetition priming in music. *J. Exp. Psychol. Hum. Percept. Perform.* **34**: 693–707.
7. Jones, D. & R. Baraniuk. 1995. An adaptive optimal-kernel time-frequency representation. *IEEE Trans. Signal Process.* **43**: 2361–2371.