35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Automated Estimation of Relative Fundamental Frequency *

Yu-An S. Lien and Cara E. Stepp

Abstract— **Relative fundamental frequency (RFF), defined as the normalized fundamental frequencies of vowels surrounding voiceless consonants, has been shown to have a characteristic pattern in healthy voices that differs from those with disordered voices (e.g. vocal hyperfunction, Parkinson's disease). However no large-scale clinical study has been performed, mainly because the current estimation protocol requires trained technicians to manually perform this time-consuming task. In this study, we developed a method to automate RFF estimation and tested the algorithm on recordings from 12 healthy participants and 12 participants with Parkinson's disease. The means and variations of RFFs estimated using the automation algorithm were similar to the 'gold standard' estimates developed by two trained technicians. The mean squared error for the automated estimates, when compared to the 'gold standard' RFF estimates, were similar to those estimated manually by an additional trained technician. Future work will focus on improving vocal cycle detection and extending the automation to estimate RFF from instances in running speech.**

I. Introduction

Relative fundamental frequency (RFF) is defined as the normalized (relative) fundamental frequencies of the vowels surrounding a voiceless consonant and can be measured from instances of speech with a vowel followed by a voiceless consonant and another vowel (VCV). RFF is an acoustic measure with a characteristic pattern that differs between healthy individuals and individuals with disordered voices (e.g. Parkinson's disease [1], and vocal hyperfunction [2, 3]). The fundamental frequencies (f_0) of ten vocal cycles before (offset of the first vowel) and after (onset of the second vowel) the consonant are normalized in semitones relative to the more steady-state portions of the vowels to account for individual pitch differences. The cycles preceding the consonant are normalized by the f_0 of the first cycle (furthest away from the consonant) of the first vowel (vowel offset) and the cycles following the consonant are normalized by the f_0 of the tenth cycle of the second vowel (vowel onset).

Healthy young individuals tend to have stable or slightly increasing offset RFF as voicing transitions into the consonant and rapidly decreasing onset RFF [4-6] as voicing transitions out of the consonant. Individuals with voice disorders tend to have lowered onset and offset RFF compared to healthy individuals [1, 2]. Additionally, in individuals with Parkinson's disease, RFF tends to be lower for those off medication than those on medication [1]. Although these results suggest promise for the use of RFF in clinical diagnosis and assessment of voice disorders, no prospective large-scale study has been initiated. In part, this is likely due to the time-consuming nature of current RFF estimation procedures.

RFF is typically estimated manually by one or two trained technicians using at most six VCV instances repeated two to six times per subject [1, 2, 7-9]. However, a recent study has shown that an increase in the number of instances used results in higher correlation with perceptual measures, and the use of six or more instances is necessary to provide a stable estimate [10]. Increasing the number of instances used will create an even larger barrier to clinical implementation of RFF. Thus, automated methods of RFF estimation are necessary to fully utilize this promising measure.

Although the mechanism underlying the observed RFF has been hypothesized as the interplay of tension [11-14], aerodynamics [4, 15, 16], and vocal fold kinematics [17, 18], the contribution of each mechanism is not clear. Elucidation of the physiological mechanisms that result in the characteristic RFF in healthy individuals and the changes that occur to these mechanisms with voice disorders are essential for clinical validation of RFF. An automated algorithm for RFF estimation will promote future work to be carried out to determine the bases of RFF.

In this paper, we will introduce a method to determine RFF automatically using selected speech samples containing a specific VCV, "ahfah" (/afa/). We tested our algorithm on speech samples from healthy young adults as well as speech samples from individuals with a disorder known to affect RFF (individuals with Parkinson's disease).

II. Methods

A. Recording procedure

The algorithm was trained on a single group of healthy young adults and then tested on two groups: healthy young

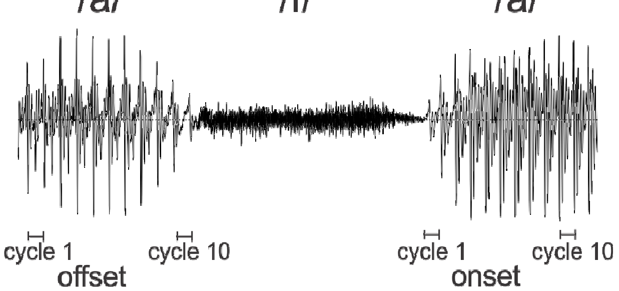

Figure 1. Acoustic recording of the phonemes /afa/. The fundamental frequencies of the offset and onset cycles can be used to estimate RFF.

*Research supported by the Dudley A. Sargent Fund of Boston University and the American Speech-Language Hearing Foundation.

Y. S. Lien is with the Biomedical Engineering Department, Boston University, Boston, MA 02215 USA (phone: 617-358-1395; fax: 617-353-5074; e-mail: slien@bu.edu).

C. E. Stepp is with the Departments of Speech, Language & Hearing Sciences and Biomedical Engineering, Boston University, Boston, MA 02215 USA (e-mail: cstepp@bu.edu).

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

adults and individuals with Parkinson's disease (PD). The training group consisted of six adults (three female, three male) between the ages 23 and 28 (MEAN = 24 years, SD = 2 years) all of whom reported no history of language, speech, or hearing disorders. The healthy young adult test group consisted of 12 adults (six female, six male) between the ages 18 and 24 (MEAN = 20 years, SD = 2 years). The PD test group consisted of 12 individuals with PD (five females, seven males) between the ages 41 and 82 (MEAN = 69 years, SD = 11 years) all of whom had been diagnosed with PD for 1.5 to 14 years, but only one of whom specifically noted voice symptoms. Each participant was instructed to repeat the VCV instance /afa/ three times in their usual, comfortable voice. All participants were native speakers of American English. Recordings were performed using a head-mounted microphone connected to a digital audio recorder sampling at 44.1 kHz and 16-bit resolution in a low-noise environment.

B. *Manual Data Analysis*

Audio files were imported into Praat [19], a software for speech analysis. The pitch range was set to 60 - 300 Hz for male recordings and 90 – 500 Hz for female recordings. Default settings were used for all other parameters. Suitability of samples for RFF analysis was determined by two trained technicians.

Onset and offset RFFs could be rejected by technicians for three reasons. First, a sample was rejected if the onset or offset contained misarticulations. Second, a sample was rejected if the first offset cycle or tenth onset cycle was not at steady state. Third, it was rejected if the phoneme was glottalized. Glottalized samples tend to have lower fundamental frequencies [20, 21] and irregular, often dicrotic, vibratory cycles [22-25], so they are not representative of typical RFF. Out of the three potential productions for each sample, the mean number of offset productions used for healthy individuals and individuals with PD were 2.9 (SD = 0.31) and 2.9 (SD = 0.28) respectively; the mean number of onset production used were 2.3 (SD = 0.47) and 2.3 (SD = 0.72) respectively.

The times (pulse timings) between ten adjacent cycles of voicing before and after the /f/ were extracted using Praat. The instantaneous fundamental frequency was calculated as the inverse of the difference between adjacent pulse timings. The instantaneous fundamental frequencies can be used to calculate the relative fundamental frequencies for all speech samples in semitones (ST) using (1) [26], in which f is the instantaneous fundamental frequency and f_{ref} is the steady state fundamental frequency.

$$ST = 39.86*\log_{10}(f / f_{ref}) \qquad (1)$$

This pulse period selection and subsequent calculations were performed collaboratively by two experienced technicians. Their RFF estimates were considered the 'gold standard'. A third experienced technician independently carried out the pulse period selection and calculation for all suitable samples. The technician's RFF estimates were used to evaluate the inter-technician precision of manual analysis.

C. *Automated Data Analysis*

Audio files were imported into MATLAB for analysis. There were three main steps for automated analysis. First, the algorithm located the RFF instance in the speech sample (RFF Instance Selection). Second, it rejected the samples based on the criteria developed in manual analysis (detailed in section B; RFF Rejection and Estimation). Lastly, the algorithm estimated the RFF.

<u>RFF Instance Selection:</u> The speech waveform was divided by the maximum absolute amplitude to normalize for sound intensity. A short-time Fourier transform was applied with a Hamming window of 20 ms and ~30% (265 samples) overlap. The signal powers for frequencies between 0 to 5158 Hz were divided into 12 bins and used to train a logistic regression model to predict the probability that a sequence window was one of the two phonemes (/a/ or /f/) or silence.

The recording always started and ended with a few seconds of silence, so the first and last five segments were assumed to be silences. The class of all other segments was determined as the one with the highest sum of the probability of the seven closest segments. Since there cannot be transitions from /f/ to silence or silence to /f/, when the program detected such a transition, it classified the point as /f/ or silence depending on the majority class of the four samples before and after the occurrence of the transition.

After all the segments were classified, the algorithm located the positions of the onset and offset vowels in the instances /afa/. The start of the offset vowel was located where silence transitions to /a/ and the end was located where the /a/ transitions into /f/. The start of the onset vowel was located where the /f/ transitions into /a/ and the end was located where the /a/ transitions into silence.

In some instances, speakers produced glottalized voicing instead of silence as they transitioned from one /afa/ to the next. When this occurred, the algorithm classified the end of the onset vowel in the preceding /afa/ and the start of the offset vowel in the subsequent /afa/ to be located about the middle of this glottalized voicing.

<u>RFF Rejection and Estimation:</u> The MATLAB algorithm was interfaced with Praat to find pulse timings that occurred within a time window specified by the MATLAB algorithm. The pitch settings in Praat were changed automatically depending on the gender of the subject. All other parameters in Praat were set to default settings. RFF was estimated using two methods (autocorrelation and cross-correlation) and averaged.

To find the offset pulse timings, the algorithm located the pulse timings in Praat between the start of the offset and the middle of the /f/. After this, the analysis window in Praat was zoomed to 25ms before and after the last sixteen pulses and the pulse timings were computed again. Similarly, to find the onset pulse timings, the algorithm located the pulse timings between the middle of the /f/ and the end of the onset. It then zoomed into an analysis window 25ms before and after the first sixteen pulse timings to find the onset pulse timings.

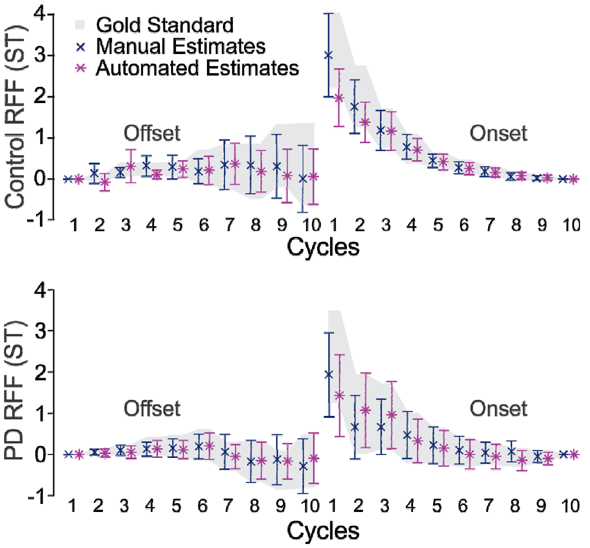

Figure 2. Top: Mean values of manual (dark navy) and automated (light pink) estimates of RFF in the control group. Bottom: Mean values of manual and automated estimates of RFF in the PD group. In both plots, the 95% confidence intervals of gold standard RFF estimates are plotted in the background. Error bars indicate 95% confidence intervals.

The criteria for rejections were similar to the ones for manual analysis. Onset and offset RFFs could be rejected for three reasons. First, a sample was rejected if the onset or offset contained fewer than 15 cycles, because this suggested that the first offset cycle or tenth onset cycle was not at steady state. Second, a sample was rejected if the variance in the pulse periods was greater than $2.9*10^{-6}$, or if the pulse periods of the nine cycles not adjacent to the consonant were 50% longer than 65th percentile of pulse periods. These criteria prevented glottalized samples which tend to have irregular or longer pulse periods (lower fundamental frequency). If the 10th offset or the 1st onset pulse period satisfied the latter criterion, then the adjacent pulse period was considered the 10th offset or 1st onset pulse period and an additional pulse period was calculated.

If the sample was usable, the differences between eleven adjacent pulse timings were taken to find the pulse periods, which subsequently were used to calculate the RFF in ST (see (1)).

F. *Evaluation*

The performance of the automation was assessed using the mean squared error (MSE) between the 'gold standard' RFF estimate and the automated estimate for each subject. Similarly, the MSE of the RFF estimated manually by the individual technician were also compared to the 'gold standard' RFF estimates.

III. Results

A. *RFF Estimation in healthy and PD populations*

Visual examinations of the means and 95% confidence intervals amongst the manual, automated, and 'gold standard' RFF estimates revealed similar trends (Fig. 2). The offset RFFs for both groups tend to be relatively stable or slightly decreasing, with somewhat higher variability in the control group. The onset RFF estimates for both groups tend to decrease as a function of cycle.

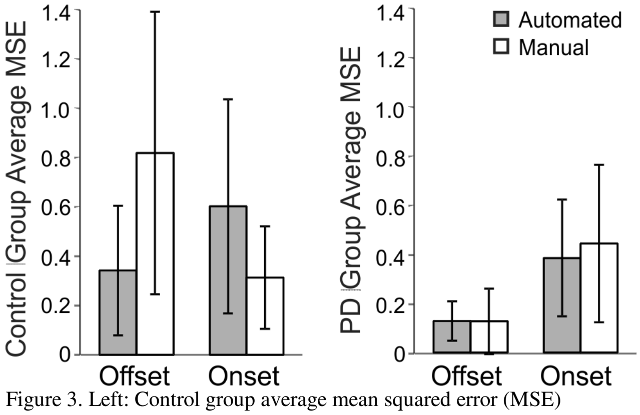

Figure 3. Left: Control group average mean squared error (MSE) relative to the gold standard. Right: PD group average MSE relative to the gold standard. Error bars indicate 95% confidence intervals.

In comparison to the gold standard, the means and 95% confidence intervals of automated RFF estimates were similar or slightly lower in most cases. The largest differences between the means were observed in offset cycles 9 and 10 and onset cycles 1 and 2 in the control group, and offset cycle 9 and onset cycle 1 in the PD group. Similarly, the means of the manual RFF estimates tend to be further from the gold standard estimates in the cycles close to the voiceless consonant.

B. *Comparison of mean squared errors between manual and automated estimates*

The MSE between the 'gold standard' RFF and the automated and independent manual estimations are shown in Figure 3. The MSE for automated RFF estimates was similar to those for the independent manual estimates in both the control group and the PD group, although in the control group the MSE varied as a function of offset/onset. In both automated and manual estimates, the MSE was higher for the control group than for the PD group.

IV. Discussion

Previous research has shown RFF to differ in individuals with disordered voices when compared to individuals with healthy voices [1-3]. RFF in healthy voices tends to be higher than in disordered voices, which is most apparent in offset cycle 10 and onset cycle 1 (the cycles closest to the voiceless consonant). These results suggest the promise of RFF for clinical diagnosis of voice disorders. However, no large scale clinical studies has been carried out, likely due to the time-consuming nature of RFF estimation. Successful development of automated methods of RFF estimation could allow for RFF to be incorporated as a feature for objective, quantitative diagnosis of voice disorders.

This paper describes an initial attempt to automate RFF estimation. The means and spreads of the RFFs estimated using the automated methods were similar to the 'gold standard' estimates provided by two expert technicians in most cases (see Fig. 2). In general, the MSE between the automated estimates and the 'gold standard' RFF were comparable to the ones between the manual estimates and the 'gold standard' RFF, indicating that a similar level of consistency can be achieved using automated estimation as is possible with manual estimation by a highly trained technician.

Although this initial work is promising, further improvements can be achieved through more advanced algorithms for automation. The current algorithm occasionally misses the last offset and first onset cycles, which resulted in the lower RFF estimates observed in Fig. 2. We expect that this problem will be resolved in the future by combining the pulse timings that result from several analysis windows in Praat, because Praat detects different vocal cycles based on the analysis windows used. The current study used an average of the standard Praat autocorrelation and cross-correlation algorithms for pitch detection: future work will compare a variety of other pitch detection algorithms to determine which are optimal for automation. In addition, instead of checking for 15 cycles, the algorithm will directly check whether the last cycle is at steady state by comparing the instantaneous frequencies of the two cycles closet to vowel mid-sections (offset cycles 1 and 2 and onset cycles 9 and 10).

Interestingly, although previous work has indicated that RFF is lowered in individuals with PD relative to those with healthy voices [1], we did not see lower RFF values in our PD group relative to our control group, whether estimated manually or automated. There are a number of potential explanations for this. The nine individuals with PD recruited in a previous study all reported symptoms of hypokinetic dysarthria, whereas the individuals in our PD sample were recruited irrespective of whether they had been diagnosed with any voice and/or speech disorder. In fact, only one of our 12 PD participants reported having voice or speech problems. Another potential reason for the similarity between our PD and control group data could be the difference in RFF stimuli used in our study relative to previous work. In our experiment, RFF was measured from non-speech samples, whereas in previous studies, RFF was measured from running speech. Our future work will extend our automation algorithms to be able to determine RFF in more complex speech samples (e.g. in running speech). This will promote additional studies to be carried out to determine the effects of speech context on RFF.

Acknowledgment

We would like to thank Caitlin I. Gattuccio, Margaux Canevari, Gabby Hands for their help with data collection, Jake Hermann for his help with the interface between MATLAB and Praat, and Emma Billiard and Lauren Kalfin for their assistance with manual RFF analysis.

References

[1] A. M. Goberman and M. Blomgren, "Fundamental frequency change during offset and onset of voicing in individuals with Parkinson disease," *J Voice,* vol. 22, pp. 178-91, Mar 2008.

[2] C. E. Stepp, R. E. Hillman, and J. T. Heaton, "The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset," *J Speech Lang Hear Res,* vol. 53, pp. 1220-6, Oct 2010.

[3] C. E. Stepp, G. R. Merchant, J. T. Heaton, and R. E. Hillman, "Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction," *J Speech Lang Hear Res,* vol. 54, pp. 1260-6, Oct 2011.

[4] A. Lofqvist, L. L. Koenig, and R. S. Mcgowan, "Vocal-Tract Aerodynamics in /aCa/ Utterances - Measurements," *Speech Communication,* vol. 16, pp. 49-66, Jan 1995.

[5] A. S. House and G. Fairbanks, "The influence of consonant environment upon the secondary acoustical characteristics of vowels," *J Acoust Soc Am,* vol. 25, pp. 105-113, 1953.

[6] N. Umeda, "Influence of segmental factors on fundamental frequency in fluent speech," *J Acoust Soc Am,* vol. 70, p. 350, 1981.

[7] C. E. Stepp, D. E. Sawin, and T. L. Eadie, "The Relationship between Perception of Vocal Effort and Relative Fundamental Frequency during Voicing Offset and Onset," *J Speech Lang Hear Res,* May 21 2012.

[8] B. C. Watson, "Fundamental frequency during phonetically governed devoicing in normal young and aged speakers," *J Acoust Soc Am,* vol. 103, pp. 3642-7, Jun 1998.

[9] R. M. Arenas, P. M. Zebrowski, and J. B. Moon, "Phonetically governed voicing onset and offset in preschool children who stutter," *J Fluency Disord,* vol. 37, pp. 179-87, Sep 2012.

[10] C. E. Stepp and T. L. Eadie, "Relative fundamental frequency as an acoustic correlate of vocal effort in spasmodic dysphonia," *J Acoust Soc Am,* vol. 129, pp. 2526-2526, 2011.

[11] A. Lofqvist, T. Baer, N. S. McGarr, and R. S. Story, "The cricothyroid muscle in voicing control," *J Acoust Soc Am,* vol. 85, pp. 1314-21, Mar 1989.

[12] K. N. Stevens, "Physics of laryngeal behavior and larynx modes," *Phonetica,* vol. 34, pp. 264-79, 1977.

[13] M. Halle and K. N. Stevens, "A note on laryngeal features," *MIT Research Laboratory of Electronics Quarterly Progress Report, vol. 101* pp. 198-213, 1971.

[14] J. Ohala, "Explanations for the intrinsic pitch of vowels," in *Monthly Internal Memorandum of the Phonology Laboratoratory, UC Berkeley.(1975)" A mathematical model of speech aerodynamics," in Speech Communication. Proceedings of the Speech Communication Seminar, Stockholm, Aug*, 1973, pp. 1-3.

[15] A. Lofqvist and R. S. Mcgowan, "Influence of Consonantal Environment on Voice Source Aerodynamics," *Journal of Phonetics,* vol. 20, pp. 93-110, Jan 1992.

[16] R. J. Baken and R. F. Orlikoff, "Changes in Vocal Fundamental-Frequency at the Segmental Level - Control during Voiced Fricatives," *J Speech Hear Res,* vol. 31, pp. 207-211, Jun 1988.

[17] N. Fukui and H. Hirose, "Laryngeal adjustments in Danish voiceless obstruent production," *Annual Report of the Institute of Phonetics, University of Copenhagen,* vol. 17, pp. 61-71, 1983.

[18] P. Ladefoged, *Three areas of experimental phonetics: Stress and respiratory activity, the nature of vowel quality, units in the perception and production of speech* vol. 15: Oxford University Press, 1972.

[19] W. Boersma and D. Weenink, "Praat: doing phonetics by computer," 5.3.04 ed, 2012.

[20] H. Hollien and R. W. Wendahl, "Perceptual study of vocal fry," *J Acoust Soc Am,* vol. 43, pp. 506-9, Mar 1968.

[21] R. E. Mcglone and T. Shipp, "Some Physiologic Correlates of Vocal-Fry Phonation," *J Speech Hear Res,* vol. 14, pp. 769-&, 1971.

[22] R. B. Monsen and A. M. Engebretson, "Study of variations in the male and female glottal wave," *J Acoust Soc Am,* vol. 62, pp. 981-93, Oct 1977.

[23] P. Moore and H. Von Leden, "Dynamic variations of the vibratory pattern in the normal larynx," *Folia Phoniatr (Basel),* vol. 10, pp. 205-38, 1958.

[24] R. W. Wendahl, G. P. Moore, and H. Hollien, "Comments on Vocal Fry," *Folia Phoniatr (Basel),* vol. 15, pp. 251-5, 1963.

[25] R. L. Whitehead, D. E. Metz, and B. H. Whitehead, "Vibratory patterns of the vocal folds during pulse register phonation," *J Acoust Soc Am,* vol. 75, pp. 1293-7, Apr 1984.

[26] R. J. Baken, *Clinical Measurements of Speech and Voice*. Austin, TX, 1987.