# *F*0-induced formant measurement errors result in biased variabilities

**Wei-Rong Chen,**[a] **D. H. Whalen,**[b] **and Christine H. Shadle**
*Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511, USA*
*chenw@haskins.yale.edu, whalen@haskins.yale.edu, shadle@haskins.yale.edu*

**Abstract:** Many developmental studies attribute reduction of acoustic variability to increasing motor control. However, linear prediction-based formant measurements are known to be biased toward the nearest harmonic of *F*0, especially at high *F*0s. Thus, the amount of reported formant variability generated by changes in *F*0 is unknown. Here, 470 000 vowels were synthesized, mimicking statistics reported in four developmental studies, to estimate the proportion of formant variability that can be attributed to *F*0 bias, as well as other formant measurement errors. Results showed that the *F*0-induced formant measurements errors are large and systematic, and cannot be eliminated by a large sample size.

[BHS]
**Date Received:** March 21, 2019 **Date Accepted:** April 22, 2019

## 1. Introduction

The variability in speech is caused by many interacting factors and can never be fully explained. In statistical modeling, researchers select a finite set of factors, mainly based on current understanding of the research questions and experimental controls, and enter them into a model. Each of these factors explains a portion of the variance of the dependent variable; the residual of the model, the unexplained variance, is considered to be "noise" or the so-called "natural variability." In speech studies, the range of such natural variability has been used to indicate the precision or the degree of active control in speech production. The pioneering work of Eguchi and Hirsh (1969) investigated the production of American English (AE) vowels in young children (3–13 yr of age) and adults ($N=84$). They reported that the variabilities of all the measured acoustic parameters [including fundamental frequency (*F*0), formant frequencies (*F*1, *F*2, *F*3), and voice onset time] decrease as the talker's age increases, and this decrease in variability serves as an indication of the maturation of speech motor control. Lee *et al.* (1999) further extended this investigation by analyzing the acoustic data of nearly 500 speakers (5–18 yr of age plus adults) and reported more acoustic characteristics based on the data; their results in general corroborate the conclusions of Eguchi and Hirsh (1969). More recently, Heald and Nusbaum (2015) studied the within-talker vowel precision across days for eight adults and concluded that vowel production is extremely precise: the within-talker formant variabilities across days were less than the just noticeable formant difference) (3%–5% of formant frequency) reported by Flanagan (1955).

Acoustic measurements are noninvasive, cost-effective, and irreplaceable in studies of speech variability. However, formant frequencies are known to suffer considerably from *F*0 bias, especially when measured by linear predictive coding (LPC) analysis (Klatt, 1986; Vallabha and Tuller, 2002; Fulop, 2010; Shadle *et al.*, 2016). Here we use the term "harmonic attraction" to emphasize the strong bias toward the closet harmonic in formant estimation, which is more serious in high-pitched speech due to time-domain aliasing and increased spacing between harmonics. This was first described by Atal (1975), but in the decades since, LPC-based analysis has continued to be the most commonly used formant measurement method in speech studies. Vallabha and Tuller (2002) explained in detail that the errors intrinsic to LPC are systematic, not random, but because *F*0 will vary widely in real speech, "it is likely that [*F*0] quantization will not distort the overall distribution of formants"; this assertion has never been tested. In order to test this hypothesis, we did a large-scale simulation of vowels, analyzed them with LPC, and compared the LPC formant estimates to the

[a] Author to whom correspondence should be addressed.
[b] Also at: Program in Speech-Language-Hearing Sciences, City University of New York, New York, New York 10016, USA.

actual formants to test whether biases remain even with large samples. This allowed us to report the degree to which the measurement errors and, for the first time, variabilities of formant frequencies derived from LPC analysis can be explained by $F0$ variability.

## 2. Methods

To model patterns that resemble those in the literature, we synthesized multiple vowels based on four studies (see Sec. 2.1). The formant frequencies were fixed at specific values and the fundamental frequency ($F0$) was varied over a realistic range [i.e., one that matched the mean and standard deviation (SD) of the previous studies]. Despite the possibility that the voice source may vary with age and gender (Iseli *et al.*, 2006; Yen-Liang and Iseli, 2008), we used a constant modal voice source for synthesis. Because the true, designated formant frequencies were known, deviations from those values can be attributed solely to the biases induced by $F0$ variability.

### *2.1 References for formant and fundamental frequency values*

We selected four sources in the literature to provide the distributions of fundamental and formant frequencies for AE vowels, which included 1, 6, 7, or 10 monophthongs. Speakers ranged in age from 3 to 18 yr old plus adults, yielding a total of 470 different combinations of reference $F0$ variabilities and formant values, roughly resembling vowel variabilities over a life span; we call each combination a reference group. The current study demonstrates how LPC fails in formant estimation using these reference values; it does not have any implications for the previously reported results. That is, we cannot retroactively correct the original results underlying the data presented in these sources. The four sources were the following:

(1) Eguchi and Hirsh (1969) (abbreviated as EH69) reported $F0$, $F1$, $F2$ values and C-to-V durations of vowel productions from 84 speakers, ranging from 3 to 13 yr of age, plus adults; gender was balanced for groups older than 10 yr of age but was not specified for those younger than 10 yr old. They produced six AE vowels / i, æ, u, ɛ, ɑ, ɔ / in two sentences, "He has a blue pen," and "I am tall" (five tokens of each), providing us with 96 reference groups (16 age-sex groups × 6 vowels). They reported both within- and between-speaker variabilities but only the within-speaker variability is included in our analysis.

(2) Lee *et al.* (1999) reported various acoustic characteristics (including $F0$, $F1$, $F2$, and $F3$), of 10 AE vowels / i, æ, u, ɛ, ɑ, ɔ, ɪ, ʌ, ʊ, ɝ / produced by 492 speakers, ranging from 5 to 18 yr of age, plus adults; gender varied across groups but was roughly balanced, providing 300 reference groups (15 ages × 2 sexes × 10 vowels). The number of repetitions per vowel for each speaker was not reported but can be inferred from the description to be 1 to 2 times. Only the values of between-speaker variabilities are available in Lee *et al.* (1999).

(3) Huber *et al.* (1999) reported $F0$ and the first three formant values of the AE vowel /ɑ/ from 180 speakers in 9 age groups, ranging from 4 to 18 yr of age, plus adults (10 female and 10 male in each age group), providing 18 reference groups (9 ages × 2 sexes × 1 vowel). The vowel was repeated 3 times by each speaker. Only the between-speaker variabilities were reported.

(4) Heald and Nusbaum (2015) (abbreviated as HN15) reported $F0$ and $F1$–$F3$ of seven vowels / i, æ, u, ɛ, ɑ, ɪ, ʌ / from eight AE adult speakers (5 female), providing 56 reference groups (8 speakers × 7 vowels). Each vowel was produced 90 times across 5 days by each speaker. Only the within-speaker variabilities were reported.

### *2.2 Vowel synthesis*

The four sources together provide distributional parameters for $F0$ and formant frequencies in real speech for 470 reference groups. For each group, we synthesized 1000 vowels (duration = 200 ms; sampling rate = 10 kHz) where the $F0$'s were randomly sampled from a distribution based on the reported $F0$ mean and SD for that reference group, and the first three formants were held constant at the reported mean values. The global $F0$ range across all synthesized vowels was 70-479 Hz. For the stimuli based on EH69, in which only $F1$ and $F2$ were reported, $F3$ was set at 1000 Hz above $F2$. The synthesis was done with Praat's (Boersma and Weenink, 2018) implementation of the Klatt synthesizer (Klatt and Klatt, 1990). Excitation sources were modal voiced glottal pulse signals (Rosenberg, 1971), and the filters were all-pole linear models. Formant bandwidths were set at the default values in Praat (i.e., $B1 = 60$, $B2 = 110$, and $B3 = 160$ Hz). In all, 470 000 stimuli were synthesized.

### *2.3 LPC formant analysis*

For each stimulus, the first three formant frequencies were estimated using LPC (Burg algorithm) (45 ms Gaussian-shaped windows, the Praat default; 2 ms steps; 10 poles) and followed by Viterbi tracking in Praat. We implemented a “seeding” method where the true formant values were provided as references during formant tracking.

Note that there are two kinds of formant measurement errors: (1) selection error, which refers to formants being missed, which will result in formant jumps and large formant error (e.g., missing true $F2$ and selecting $F3$ as $F2$); and (2) estimation error, which refers to the measurement error given that the correct spectral prominence in the vicinity of the true formant is selected. The focus of the current paper, $F0$-induced error, is related only to the latter. Therefore, our seeding method very nearly guarantees that the measurement errors in this study came from estimation errors, but not from selection error. This method also simulates the errors that occur when humans correct selection errors by visual inspection of the LPC spectrum.

### *2.4 Measurements of* F0*-induced biases*

We report three measurements of $F0$-induced biases in this study: (1) token error, (2) central tendency error, and (3) formant variability.

(1) Token error is calculated as $|e_i| = |\hat{f}_i - f|$, where $\hat{f}_i$ is the LPC-estimated formant frequency for the $i$th sample and $f$ is the true formant frequency.

(2) Central tendency error is calculated by subtracting the true formant frequency $f$ from the mean of multiple formant estimations: $\bar{e} = [(1/n) \cdot \Sigma \hat{f}_i] - f$, where $n$ is 1000, the number of samples for each group. Positive error values indicate over-estimation of formant frequency over multiple measurements; negative values, under-estimation. *If token errors are random, then central tendency error is expected to be zero.* This calculation is used to simulate the error of the estimated central tendency when a large number of tokens are averaged. For each sample, we also calculated the Boolean value denoting whether the nearest harmonic is on the same side of the true formant as the estimated formant $\hat{f}_i$, as $b_i = u(hf_i - f)$ *XNOR* $u(\hat{f}_i - f)$, where $hf_i$ is the harmonic nearest to $f$ for the $i$th sample, $u(x)$ is the unit step function [$u(x)$ is equal to 1 if $x \geq 0$, and is zero elsewhere]. *XNOR* is the exclusive-nor operation; if both operands are greater than or both are less than zero, $b_i = 1$; otherwise, $b_i = 0$. The mean $\bar{b}_i = (1/n) \cdot \Sigma b_i$ denotes the proportion of error tokens in each group that can be explained by harmonic attraction.

(3) Formant variability is calculated by dividing the SD of the measured formant frequencies for each age-gender-vowel group by the corresponding group mean. This ratio is known as *relative standard deviation* (RSD), or *coefficient of variation*, and has been used in Eguchi and Hirsh (1969) and Lee *et al.* (1999) to indicate normalized variabilities.

## 3. Results

### *3.1 Degree of formant measurement error*

Figure 1 presents the probability distributions of the absolute token measurement errors $|e_i|$ binned into four different $F0$ ranges: (1) 70–100 Hz, (2) 100–200 Hz, (3)

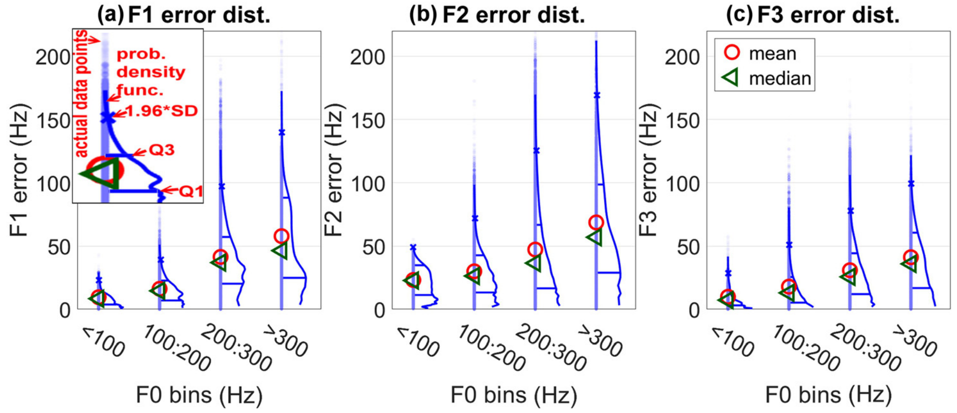

Fig. 1. (Color online) The results of token errors (see Sec. 2.4) grouped in four $F0$ bins for (a) $F1$, (b) $F2$, and (c) $F3$ frequencies. Each curvy line, circle, triangle, and cross marker indicate, respectively, the probability density estimate, mean, median, and 1.96*SD boundary for each $F0$ bin.

200–300 Hz, and (4) 300–429 Hz. The probability density function for each bin was estimated by nonparametric kernel distribution (kernel density estimation, using "*ksdensity*" function in MATLAB), cutoff at the probability of 1%. In general, formant measurement error increases as *F*0 increases, and when *F*0 is higher than about 200 Hz, the error increases substantially. The *F*0-induced errors in *F*3 measurement are lower, presumably due to a wider formant bandwidth that encompasses more harmonics (in favor of LPC). *F*2 errors are comparatively less serious than *F*1 errors but the distributions indicate that when *F*0 is higher than 300 Hz, there were occasionally *F*2 errors greater than 150 Hz (which is not negligible for *F*2). Among the first three formants, *F*1 measurement is the most affected by *F*0-bias. When *F*0 is above 200 Hz, the median error for *F*1 measurement is approximately 40–50 Hz. When *F*0 is above 300 Hz, the third quartile of *F*1 error distribution is approximately 90 Hz. (Breaking down by vowel, errors expressed as percent of error relative to formant value are 8% for /ɑ/, 23% for /u/, 18% for /i/, and 10% for /æ/.)

### *3.2 Direction of formant measurement error*

Figure 2 displays the central tendency errors $\bar{e}$ in the estimation of central tendencies from repeated measurements for *F*1. Each datapoint in Fig. 2 represents the central tendency error for each age-gender-vowel group: the true *F*1 frequency is subtracted from the mean of 1000 formant measurements. The proportions of harmonic explanation $\bar{b}_i$ grouped into four proportion ranges, are represented by four different symbols in Fig. 2 (see the legend).

If the hypothesis is that *F*0-induced formant estimation error is random (and therefore the individual errors cancel each other out), then we would expect that the central tendency errors displayed in Fig. 2 should cluster around zero. However, the results are inconsistent with this hypothesis—the errors for nine out of ten vowels (except /ɝ/) show clustering patterns either above or below zero. *F*1 is mostly overestimated for the high vowels /i/ and /u/, and underestimated for the low vowels /ɑ/ and /ɔ/. This is due to the fact that the harmonic closest to *F*1 is on average above *F*1 for /i/ and /u/, and below *F*1 for /ɑ/ and /ɔ/. The errors for *F*2 and *F*3 showed similar patterns: most vowels were either over- or underestimated (with similar scale as for *F*1 errors), except for the *F*2 errors of /i/ which clustered around zero. As can be seen in Fig. 2, the dominant proportion of harmonic attraction is in the range of 75%–100% (circles) for 433 out of 470 error tokens. The average proportion of errors explained by harmonic attraction was 89% for *F*1, 77% for *F*2, and 66% for *F*3. That is, even when

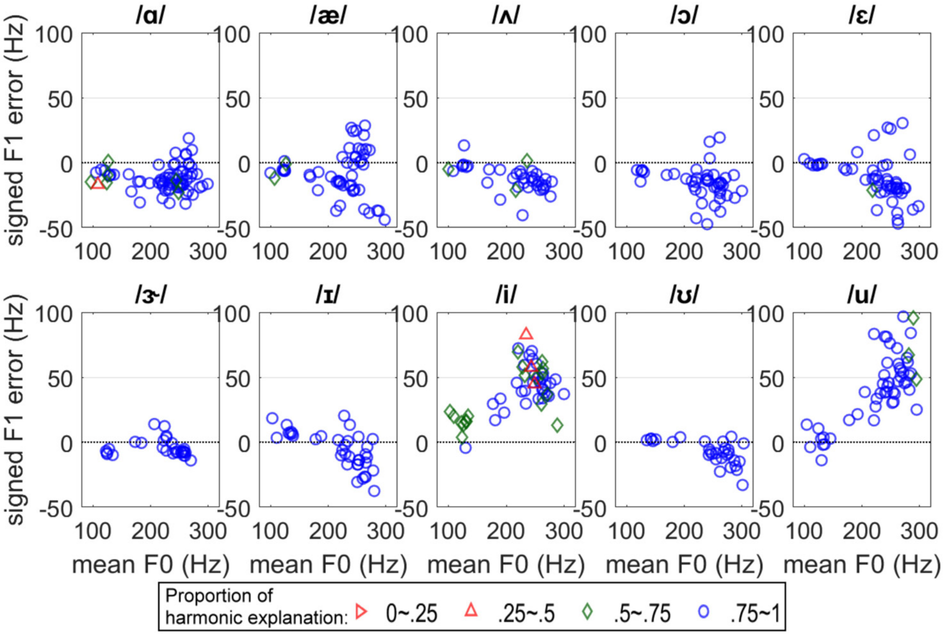

Fig. 2. (Color online) Central tendency errors plotted against *F*0 for each vowel. Each datapoint represents the mean of 1000 formant measurements minus true *F*1 for each group, plotted against mean *F*0 for that age-gender-vowel group. Positive values indicate over-estimation of formant frequency. The four different symbols indicate four ranges of proportions of error tokens that can be explained by harmonic attraction. Here 433 of the total 470 symbols are plotted in circles.

the measurement is relatively accurate, it is just as likely to be because a harmonic was near the true formant as that the LPC analysis *per se* was correct.

### *3.3* F0*-biased formant variability*

Last, we present the formant variability caused by *F*0-induced formant measurement errors. Figure 3(a) summarizes the within-speaker *F*1 variabilities in RSD for each age-gender-vowel group, and Fig. 3(b) the corresponding *F*0 range for each group (error bars represent ±1.96 * mean within-speaker SD) in EH69 and HN15. Figures 3(a) and 3(b) largely mimic Figs. 7 and 9 of EH69, respectively, since HN15 only reported values for adults. The simulated (*F*0-induced) *F*1 variabilities for the same groups are shown in Fig. 3(c) (note that all the true formant variabilities in our simulations were zero). It can be seen that the simulated (*F*0-biased) *F*1 variabilities have the same trend as in Fig. 3(a): Formant variability decreases as age increases. This is due to the fact that younger groups have higher *F*0s that induce larger formant measurement errors and thus higher formant variabilities. Figure 3(d) shows the ratios of the simulated variabilities [Fig. 3(c)] to the reported variabilities [Fig. 3(a)]; these ratios indicate the degree of the expected inflation of formant variabilities due to LPC analysis, assuming the reported variabilities in Fig. 1(a) are the ground truths. The average of these ratios across groups was 0.76 for *F*1 and 0.58 for *F*2, indicating that more than half of the reported *F*1 and *F*2 variabilities are likely due to *F*0 variability. To further explain this, we ran four separate sets of simulations (each with 1000 tokens) of the vowel /i/ (the *F*0 and *F*1 were centered at 220 and 400 Hz, respectively) with the same *F*0 variability (SD = 5 Hz) but with four different underlying *F*1 variabilities (SD = 16, 10, 5, and 3 Hz). The results showed that the LPC-measured *F*1 variabilities (SDs) for all four sets were approximately 10 Hz in all cases. This agrees with the variability of the second harmonic, which is the harmonic closest to true *F*1 (see Fig. 4).

## 4. Discussion

### *4.1 Confounding contribution of* F0 *to formant measurement error*

In this study, we have reported *F*0 and formant combinations matched to published reports, which show the scale and range of formant measurement errors intrinsic to LPC. Our results showed that the overall LPC errors increase dramatically as *F*0 increases, particularly above 200 Hz. In accord with previous studies (Klatt, 1986; Vallabha and Tuller, 2002; Fulop, 2010; Shadle *et al.*, 2016), *F*1 is the formant most affected, especially for *F*0 above 200 Hz; the mean *F*1 error is 41 Hz (7%) when *F*0 is between 200 and 300 Hz and is 58 Hz (9%) for *F*0 above 300 Hz. Even when *F*0 is lower, between 100 and 200 Hz, the mean *F*1 error (16 Hz) is higher than the perceptual threshold of *F*1 change of 14 Hz (Kewley-Port and Watson, 1994). *F*2 measurement is also considerably influenced by *F*0 bias when *F*0 is above 300 Hz (which is not uncommon in children's speech) with a mean error around 70 Hz and the third quartile approximately 100 Hz. These scales and ranges of errors reported here indicate that errors due to harmonic attraction in LPC formant measurements are substantial, even when human correction is involved (our seeding method simulates human correction).

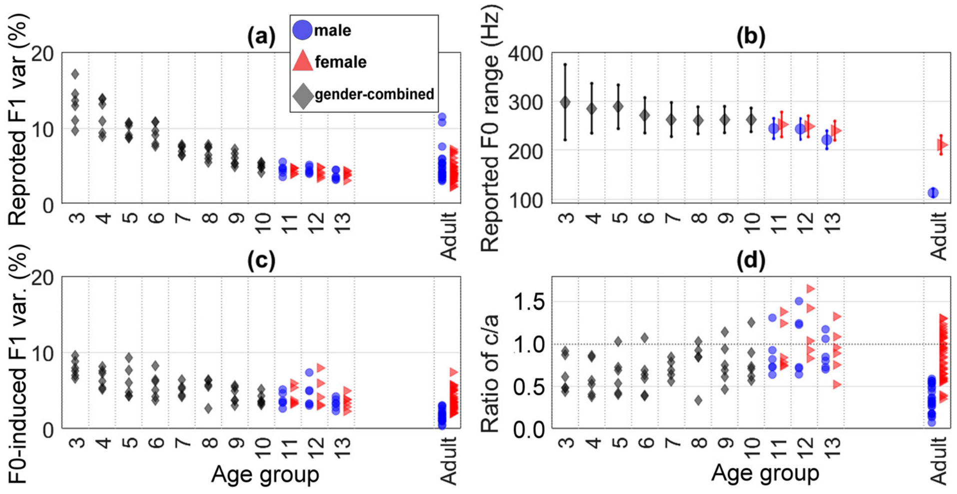

Fig. 3. (Color online) (a) The *F*1 variabilities reported in EH69 and HN15. (b) The reported *F*0 ranges in EH69 and HN15. The error bars represent ±1.96 * mean within-speaker SD. (c) Our simulated *F*1 variabilities when *F*1 is held constant and *F*0 varied. (d) The ratios of the simulated variabilities to the reported variabilities.

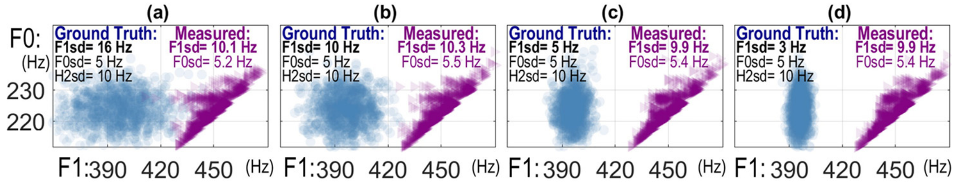

Fig. 4. (Color online) Four sets of simulations of the vowel /i/ with the same $F0$ variability (SD = 5 Hz) but four different underlying $F1$ variabilities (from left to right, the true $F1$ SD = 16, 10, 5, and 3 Hz); light (blue) circles indicate the ground truth and dark (purple) triangles the LPC measured $F1$ values.

Note that our simulations were as favorable as possible to the LPC algorithm (i.e., no noise, optimal parameter setting, ground-truth provided in formant tracking, and all-pole synthesis); therefore, formant measurement errors in natural speech are expected to be much larger than reported here.

### *4.2 Vowel-specific formant measurement error*

If errors are random, they will average out to show a genuine central tendency if measured with a large sample size. Given the condition that multiple samples are taken where the true formant $f$ is stable and $F0$ is varied such that the nearest harmonic $hf_i$ moves symmetrically through $f$, $e_i$ will cancel each other out and $\bar{e}$ will be close to zero. Our simulation with realistic $F0$ and formant ranges shows that the first three formant measurements for nearly all vowels do *not* conform to such conditions, and as a result their central tendency errors systematically clustered either above or below zero (except for $F1$ of /ɝ/ and $F2$ of /i/). Not surprisingly, some vowels are almost always overestimated while others are underestimated in formant measurements. This indicates that the measurement error is not random but systematically conditioned by vowel-specific $F0$ and formant ranges. One direct consequence of this is that the vowel space area (VSA) for children will be underestimated, as low vowels appear higher than they really are (lower $F1$), and high vowels appear lower (higher $F1$). For example, when $F0$ is 234 Hz, the LPC-estimated VSA is reduced to 85% of the true VSA, calculated from the data in Fig. 2. Thus it may be that the previously reported shrinking of the vowel space in development (see review in Kent and Vorperian, 2018) is, in fact, greater than had been thought.

### *4.3 Formant variability explained by* F0 *variability*

While there is a consensus in the literature that the maturation of speech motor control reduces variability in articulatory and acoustic measures of speech production, higher $F0$ in younger speakers results in greater LPC-measurement errors, thus increasing variability in formant frequencies. Computation of the ratio of simulated formant variability to the reported formant variability, as described above, revealed the degree of potential inflation of formant variability when measured by an LPC-based method. Again, such inflation of formant variability cannot be reduced by using a large sample size and persists in spite of human correction of formant tracks.

## 5. Final remarks on formant measurement

We have reported the scale of formant measurement errors and inflation of formant variability in a wide range of $F0$ and formant combinations, mostly due to the $F0$-bias intrinsic to LPC-based formant analysis. Recent developments in weighted linear prediction-based methods (e.g., Alku *et al.*, 2013; Airaksinen *et al.*, 2014; Gowda *et al.*, 2017) have been proposed specifically to solve this issue, but the requirement of glottal pulse information (e.g., EGG data or accurate epoch-detection algorithm) restricts its application. Manual formant measurement of the reassigned spectrogram (RS) (Rihaczek, 1968; Nelson, 2001; Fulop and Fitz, 2006) has been reported to be the most accurate and resistant to $F0$ bias (Shadle *et al.*, 2016), but automatic methods based on RS do not currently exist. More work in this direction is needed.

**Acknowledgments**

Portions of this work were presented as 4aSC29 at the 176th Meeting of the Acoustical Society of America, Victoria, BC, Canada (November 2018). This research was supported by Grant No. NIH DC-002717 to Haskins Laboratories. We thank two anonymous reviewers for helpful comments.

## References and links

Airaksinen, M., Raitio, T., Story, B., and Alku, P. (**2014**). "Quasi closed phase glottal inverse filtering analysis with weighted linear prediction," IEEE/ACM Trans. Audio, Speech, Lang. Process. **22**, 596–607.

Alku, P., Pohjalainen, J., Vainio, M., Laukkanen, A.-M., and Story, B. H. (**2013**). "Formant frequency estimation of high-pitched vowels using weighted linear prediction," J. Acoust. Soc. Am. **134**, 1295–1313.

Atal, B. S. (**1975**). "Linear prediction of speech—Recent advances with applications to speech analysis," in *Speech Recognition*, edited by D. R. Reddy (Academic Press, New York), pp. 221–230.

Boersma, P., and Weenink, D. (**2018**). "Praat: Doing phonetics by computer (version 6.0.37) [computer program]," http://www.praat.org (Last viewed February 20, 2018).

Eguchi, S., and Hirsh, I. J. (**1969**). "Development of speech sounds in children," Acta Oto-laryngol. Suppl. **257**, 1–51.

Flanagan, J. L. (**1955**). "A difference limen for vowel formant frequency," J. Acoust. Soc. Am. **27**, 613–617.

Fulop, S. A. (**2010**). "Accuracy of formant measurement for synthesized vowels using the reassigned spectrogram and comparison with linear prediction," J. Acoust. Soc. Am. **127**, 2114–2117.

Fulop, S. A., and Fitz, K. (**2006**). "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," J. Acoust. Soc. Am. **119**, 360–371.

Gowda, D., Airaksinen, M., and Alku, P. (**2017**). "Quasi-closed phase forward-backward linear prediction analysis of speech for accurate formant detection and estimation," J. Acoust. Soc. Am. **142**, 1542–1553.

Heald, S. L. M., and Nusbaum, H. C. (**2015**). "Variability in vowel production within and between days," PLoS One **10**, e0136791.

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., and Johnson, K. (**1999**). "Formants of children, women, and men: The effects of vocal intensity variation," J. Acoust. Soc. Am. **106**, 1532–1542.

Iseli, M., Yen-Liang, S., and Alwan, A. (**2006**). "Age- and gender-dependent analysis of voice source characteristics," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, p. I–I.

Kent, R. D., and Vorperian, H. K. (**2018**). "Static measurements of vowel formant frequencies and bandwidths: A review," J. Commun. Disorders **74**, 74–97.

Kewley-Port, D., and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.

Klatt, D. H. (**1986**). "Representation of the first formant in speech recognition and LF models of the auditory periphery," in *Montreal Satellite Symposium on Speech Recognition, 12th International Congress on Acoustics*, edited by P. Mermelstein (Toronto), pp. 5–7.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.

Nelson, D. J. (**2001**). "Cross-spectral methods for processing speech," J. Acoust. Soc. Am. **110**, 2575–2592.

Rihaczek, A. (**1968**). "Signal energy distribution in time and frequency," IEEE Trans. Inf. Theory **14**, 369–374.

Rosenberg, A. E. (**1971**). "Effect of glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Am. **49**, 583–590.

Shadle, C. H., Nam, H., and Whalen, D. H. (**2016**). "Comparing measurement errors for formants in synthetic and natural vowels," J. Acoust. Soc. Am. **139**, 713–727.

Vallabha, G. K., and Tuller, B. (**2002**). "Systematic errors in the formant analysis of steady-state vowels," Speech Commun. **38**, 141–160.

Yen-Liang, S., and Iseli, M. (**2008**). "The role of voice source measures on automatic gender classification," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4493–4496.