INTRODUCTION

Fundamental frequency \( (f_o) \) is one of the key parameters used for the quantitative description of voice signals.\(^1\)\(^-\)\(^5\) \( f_o \) represents the rate of vibration of the laryngeal sound generator, typically consisting of the vocal folds in humans and most mammals. \( f_o \) detection is performed under the assumption that the analyzed sound source exhibits periodic vibration.

A time series such as the (acoustic) voice signal is said to be periodic when it precisely repeats itself at certain intervals, mathematically expressed as

\[
x(t ± nT_o) = x(t)
\]

where \( t \) is time, \( n \) is a positive integer, and \( T_o \) is the period,\(^6\) that is, the duration of one glottal cycle. The smallest possible value of \( T_o \) of a periodic time series that satisfies Equation 1 is called the fundamental period of that time series, and its inverse is the fundamental frequency:

\[
f_o = \frac{1}{T_o}
\]

Several different ways of denoting fundamental frequency are used in the literature (eg, \( F0 \) or \( f_0 \) with a subscript zero). However, a recent consensus paper suggests to use the denotation \( f_o \) with a lower letter \( f \) and the character \( o \) (for “oscillatory”) instead of the zero (for “zero harmonic”) as a subscript.\(^7\)

\( f_o \) is often confused with “pitch.” \( f_o \) is a property of the vibration of a physical system, measured in hertz [Hz]. In contrast, pitch is a psychoacoustic quantity, defined as “that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high.”\(^8\) In quite a few cases the two quantities approximate each other, but not always. Hence, the term “pitch” should only be used if (human) perception is addressed, and be avoided when laryngeal sound generation is described as a physical system.

Voice, as practically any other biosignal, is never purely periodic. Rather, it is nearly periodic at best (some authors use the term “quasi-periodic,” which, however, is reserved for describing a signal with two individual fundamental frequencies\(^6\)\(^,\)\(^9\)). For one, \( f_o \) traces typically contain linear or quadratic terms, introduced by gradual changes of \( f_o \). Additionally, even the most steady vocalizations contain slight cycle-to-cycle alterations—see Ref. 9 for a very good discussion. More severe phenomena are constituted by irregularity/chaos, subharmonics (“period doubling,” “period tripling,” etc) or multiphonia or biphonation, constituted by two independent sound sources.\(^1\)\(^,\)^\(^12\) These issues make \( f_o \) detection nontrivial, particularly so in pathologic voices, certain singing styles, and in animal bioacoustics, where often the laryngeal sound source exhibits nonlinear phenomena like irregularity, subharmonics, and bifurcations between different vibratory states.\(^13\)

Strictly speaking, \( f_o \) can thus not be calculated for voice signals, because \( f_o \) is a property of purely periodic signals. Consequently, there is always a certain degree of inherent inaccuracy in any \( f_o \) estimation. In the words of Owren and Linker (1995), “All pitch extraction techniques are found to fail under some circumstances, which places a burden on the investigator to consistently monitor the performance of...
each routine being used. Regrettably, apart from some informal recommendations, no rigorously established limit or respective error ranges for the acceptable degree of irregularity have been established. This makes comparison of \( f_o \) data ranges presented in different studies highly problematic.

One additional complication of \( f_o \) detection is sometimes introduced by the degeneration of the analyzed acoustic signal by background noise. Lacking anechoic chambers or other adequately sound-treated rooms, in a medical setting this problem can be circumvented by directly assessing the process of laryngeal sound production, for example, via the glottal area waveform, derived by analysis of endoscopic laryngeal high-speed videos. However, the respective equipment is expensive and not always available.

A noninvasive alternative for assessing laryngeal vibration is electroglottography (EGG), pioneered by Fabre in 1957. In EGG, a high-frequency, low-voltage current is passed between two electrodes, which are placed on either side of the thyroid cartilage. Changes in vocal fold contact area during vocal fold vibration result in admittance variations, and the resulting EGG signal is proportional to the relative vocal fold contact area. A number of parameters quantitatively describing the laryngeal sound generation process can be extracted from a properly recorded EGG signal. Among others, the EGG signal is an ideal candidate for assessment of the (time-varying) \( f_o \) because it is neither influenced by vocal tract acoustics nor by background noise.

Even under optimal conditions there can be a certain degree of distortion in an acquired EGG signal—see, for example, Ref. 20 for a discussion. Further quality degeneration of the EGG signal can be introduced by inadequately positioned EGG electrodes (eg, caused by excessive larynx or neck movement); reduced conductivity between EGG electrodes (eg, caused by excessive larynx position); electrical interference due to tissue fat, beards, or fur; or noise introduced by the degeneration of the analyzed acoustic signal. The resultant information was used to drive a kinematic vocal fold vibration model. The model's default parameters were used \((Q_s = 0.3; Q_b = 3.0; Q_p = 1.0; Q_{0.3} = 0.2)\). This process resulted in synthetic EGG signals with nonlinearly increasing \( f_o \), as illustrated in Figure 1. The time offset and the period of the resulting glottal cycles within each synthesized signal were stored for later comparison with the analysis results from the tested algorithms.

As mentioned in the introduction, there are a number of factors that can introduce distortions into the recorded EGG signal and make \( f_o \) estimation problematic. To test the potential effect of these factors, the following features were introduced into the synthesized EGG signals at various degrees:

1. **Random \( f_o \) variation**: When generating the individual EGG cycles, their respective period was allowed to vary randomly within a certain range. This processing step was introduced after sorting the \( f_o \) values retrieved from the Gaussian distribution (see previous discussion). The final \( f_o \) of consecutive cycles within each synthesized signal was determined by

\[
f'_o(t) = f_o [1 + \alpha (RND[0..1] - 0.5)]
\]

where \( \alpha \) is the \( f_o \) random factor, which was varied between 0 (no \( f_o \) variation) and 0.3. A comparison between \( \alpha \) and the relative average perturbation (RAP), a voice quality parameter to assess pathologic human voice production, suggests a relationship of \( RAP = 0.2118 \alpha + 0.0029 \), \( R_\alpha = 0.9996 \). (The y-intercept of 0.0029 was introduced by the nonlinear increase of \( f_o \) in the synthesized signals.) As a reference, for healthy humans, RAP values of 0.0021–0.0089 were reported, which would be the equivalent of \( \alpha = [-0.0038...0.0283] \). Pathologic voices were measured to have RAP values of 0.0068–0.0452, corresponding to \( \alpha = [0.0187...0.1997] \).

2. **Subharmonics**: The presence of subharmonics, a relatively common feature in mammalian vocalization, was simulated by scaling the amplitude of every other synthesized EGG glottal cycle by \((1 - \beta)\), where the factor \( \beta \) was varied between 0 and 0.3. Nonzero values of \( \beta \) resulted in the appearance of period-2 subharmonics (period doubling). The parameter value range follows Bergan and Titze, who found that the perceptual pitch-drop of an octave occurred at amplitude modulation rates of 10%–30%.

3. **Amplitude drift**: The temporal variation of the EGG signal amplitude was simulated by introducing a sinusoidally varying amplitude modulation at an arbitrarily high modulation rate.

**MATERIALS AND METHODS**

**Synthetic test signals**

A set of synthesized EGG signals at various stages of corruption were generated at a sampling frequency of 48,000 Hz. Each synthesized signal had a duration of 2 seconds. The \( f_o \) information for each glottal cycle within a synthesized signal was derived randomly from a Gaussian distribution centered around 1000 Hz with a standard deviation of 500 Hz. Only \( f_o \) data between 100 Hz and 2000 Hz were considered. This extended range was chosen to encompass the singing voice range of humans and vocalizations of some nonhuman mammals.

The \( f_o \) values were sorted in ascending order, and the resulting information was used to drive a kinematic vocal fold vibration model. The model's default parameters were used \((Q_s = 0.3; Q_b = 3.0; Q_p = 1.0; Q_{0.3} = 0.2)\). This process resulted in synthetic EGG signals with nonlinearly increasing \( f_o \), as illustrated in Figure 1. The time offset and the period of the resulting glottal cycles within each synthesized signal were stored for later comparison with the analysis results from the tested algorithms.

As mentioned in the introduction, there are a number of factors that can introduce distortions into the recorded EGG signal and make \( f_o \) estimation problematic. To test the potential effect of these factors, the following features were introduced into the synthesized EGG signals at various degrees:

1. **Random \( f_o \) variation**: When generating the individual EGG cycles, their respective period was allowed to vary randomly within a certain range. This processing step was introduced after sorting the \( f_o \) values retrieved from the Gaussian distribution (see previous discussion). The final \( f_o \) of consecutive cycles within each synthesized signal was determined by

\[
f'_o(t) = f_o [1 + \alpha (RND[0..1] - 0.5)]
\]

where \( \alpha \) is the \( f_o \) random factor, which was varied between 0 (no \( f_o \) variation) and 0.3. A comparison between \( \alpha \) and the relative average perturbation (RAP), a voice quality parameter to assess pathologic human voice production, suggests a relationship of \( RAP = 0.2118 \alpha + 0.0029 \), \( R_\alpha = 0.9996 \). (The y-intercept of 0.0029 was introduced by the nonlinear increase of \( f_o \) in the synthesized signals.) As a reference, for healthy humans, RAP values of 0.0021–0.0089 were reported, which would be the equivalent of \( \alpha = [-0.0038...0.0283] \). Pathologic voices were measured to have RAP values of 0.0068–0.0452, corresponding to \( \alpha = [0.0187...0.1997] \).

2. **Subharmonics**: The presence of subharmonics, a relatively common feature in mammalian vocalization, was simulated by scaling the amplitude of every other synthesized EGG glottal cycle by \((1 - \beta)\), where the factor \( \beta \) was varied between 0 and 0.3. Nonzero values of \( \beta \) resulted in the appearance of period-2 subharmonics (period doubling). The parameter value range follows Bergan and Titze, who found that the perceptual pitch-drop of an octave occurred at amplitude modulation rates of 10%–30%.

3. **Amplitude drift**: The temporal variation of the EGG signal amplitude was simulated by introducing a sinusoidally varying amplitude modulation at an arbitrarily high modulation rate.
fixed rate of $f_{AM} = 2.27$ Hz. In particular, the EGG signal was multiplied by

$$(1 - \gamma) + \frac{\gamma [1 + \sin(2\pi f_{AM} t)]}{2}$$

where the amplitude modulation factor $\gamma$ was varied between 0 and 0.6 across synthesized signals.

4. **Mains hum**: A mains hum signal with a duration of 2 seconds was synthesized as a harmonic series at 100 Hz $f_o$ with a steadily decaying spectral envelope, using a spectral slope of $-6$ dB per octave. A total of 20 harmonics were included. This mains hum signal was scaled by the factor $\delta$ and then added to the normalized synthesized EGG. The amplitude scaling factor $\delta$ was varied between 0 (no mains hum added) and 0.3.

5. **Baseline drift**: The baseline offset of the signal was allowed to vary sinusoidally at an arbitrarily defined fixed rate. In particular, the following baseline drift was added to the synthesized signals: $0.5 \epsilon [1 + \sin(2\pi f_{BD} t)]$, where $f_{BD} = 1.71$ Hz. The factor $\epsilon$ was varied between 0 (no baseline drift) and 0.8 across the synthesized signals.

6. **Noise**: Finally, colored noise was added to the synthesized EGG signal to simulate various signal-to-noise ratios (SNR). SNR was varied in a range of $-5$ dB to 35 dB. These values were taken from a recent study which reported these surprisingly low SNR ranges for EGG signals recorded from humans in laboratory conditions.\textsuperscript{24} The noise was generated by scaling the frequency components of white noise in the frequency domain during a forward-backward Fourier transform. The amplitudes for the frequency-dependent scaling were derived from averaged noise contained in previously recorded EGG signals.\textsuperscript{24}

Deviating from the “best case” of the synthesized EGG signal, six data sets were generated, where each of the aforementioned six parameters was varied in isolated fashion at 21 equidistantly spaced intervals (see Figure 1 for an
example). Additionally, one data set was generated where all six parameters were varied at once (termed “compound” scenario in the remainder of this text). In this fashion, a total of 147 (21 × 7) synthesized EGG signals were produced.

**Evaluated algorithms**

Owing to frequently occurring linear/quadratic trends and cycle-to-cycle aberrations in a voice signal, quantitative analysis focuses on the time-varying rather than the mean \(f_o\). This is either achieved by short-term windowed approaches,\(^25\) where short portions of the voice signal are evaluated at consecutive time instants, or by estimation of the so-called glottal closure instants (GCI).\(^26\) GCI detection operates on the assumption that the major sound generation event occurs at the instant of glottal closure, that is, (partial) collision of the laryngeal or syringeal tissue, and that each glottal cycle has a period (sometimes called “epoch”\(^25\)) that is determined by two consecutive GCIs. Recalling Equation 2, the time-varying \(f_o\) is then found by taking the inverse of the period.

\(f_o\) estimators and GCI detectors may operate on different computational principles. The majority of them are rooted in either the time domain (looking at similarities or recurrent features in the voice signal) or in the frequency domain (by further analyzing the time-varying spectrum of the voice signal, as produced by a Fourier transform). Alternative approaches include, for example, wavelet analysis\(^28\–\(^30\) or phase space analysis.\(^31\) Description of the concepts involved in the various algorithms is beyond the scope of this paper. The reader is referred to the landmark textbook by Hess.\(^25\) Some key approaches are described in Owren and Linker’s summary.\(^5\) Further, good overviews are given by Talkin\(^32\) and Drugman et al,\(^33\) the latter of which discusses some more recent developments.

Overall, a surprisingly large number of \(f_o\) estimators have been described in the literature. In 1983, Hess already stated that “literally hundreds of pitch-determination methods and algorithms have been developed.”\(^25\) In Appendix S1 of this paper, we include a nonexhaustive list of 75 \(f_o\) and GCI estimators, addressing some past and recent developments, and providing web links to free source code or software applications where applicable.

Given the number of options, a practical solution had to be found for selecting the algorithms tested in this paper. Besides focusing on the algorithms included in the *Praat* software package,\(^34\) our main selection criterion was (1) free availability of the algorithm source code, and (2) ability to operate the algorithm within a free software environment, that is, the *Linux* operating system, and, where applicable, *GNU Octave,\(^35\) the free equivalent to *MATLAB*. A total of 13 such algorithms were included in this study:

- five algorithms from the *Praat* software, version 5.4.06.
- The following methods were tested in this study: “to Pitch (ac),” “to Pitch (SHS),” “to PointProcess (periodic, cc),” “to PointProcess (periodic, peaks),” and “to PointProcess (zeros).” These algorithms are referred to as *Praat* (AC), *Praat* (SHS), *Praat* (periodic cc), *Praat* (periodic peaks), and *Praat* (zeros), respectively, for the remainder of this text. Preliminary analysis suggested that *Praat's* methods “to Pitch (SPINET)” and “to PointProcess (extrema)” produced greatly inferior results. These two algorithms were thus excluded from this report;
  - the DECOM algorithm\(^36\) presented in \(^37\);
  - the DYPSA GCI algorithm, introduced by Kounoudes et al\(^38\) and described in further detail by Naylor et al\(^39\);
  - The NDF (nearly defect-free) \(f_o\) detector,\(^40\) implemented as MulticueF0v14.m, version 2016-06-30;
  - The RAPT algorithm,\(^32\) implemented as fxrapt.m in the voicebox package;
  - David Talkin’s *REAPER* algorithm (unpublished work; https://github.com/google/REAPER);
  - the SIGMA GCI detector, developed by Thomas and Naylor\(^40\);
  - the SWIPE’ algorithm, developed by Camacho and Harris\(^41\);
  - the YAGA GCI detector, developed by Thomas et al\(^42\);

Web links for downloading the software of the algorithms utilized in this comparison can be found in the supplementary materials.

All algorithms were controlled through a set of custom scripts written in Python by author C.T.H., operated on Ubuntu *Linux* 16.04 LTS. *Praat* and the compiled C-code of REAPER were accessed through command-line pipes. All other algorithms were available as *MATLAB* (Math-Works, Natick, Massachusetts) code. They were thoroughly tested in *GNU Octave* 4.0 and were then embedded into the custom *Python* code through *Python*’s oct2py wrapper module for *MATLAB/Octave* code.\(^43\) For all algorithms, the respective standard parameters were used, except for the upper and lower limits, which (where possible) were specified as 100 Hz and 2000 Hz, respectively. The upper frequency limits of REAPER and the voicebox-based DYPSA, RAPT, and SIGMA algorithms had to be changed from 500 Hz to 2000 Hz in the respective source code. All \(f_o\) detection algorithms (*Praat* (AC), *Praat* (SHS), NDF, RAPT, REAPER, SWIPE’) were operated at a time step of 1 ms.

**Combining algorithm outputs**

Preliminary assessment of the performance of the algorithms suggested that there was no single algorithm that performed best under all conditions. Rather, the SIGMA GCI detector and the *Praat* autocorrelation (AC) \(f_o\) estimator showed the most robust performance in different subsets of the synthesized data (see Results). In an attempt to consolidate the benefits of these two algorithms, a custom analysis approach (denoted as CUSTOM for the remainder of this paper) was implemented as follows: SIGMA GCI data were converted to \(f_o\) information at a time-step of 1 ms. For each data point (totaling 2000 for 2 seconds of synthesized sound), the difference between \(f_o\) data from *Praat* AC and
SIGMA was computed, expressed in octaves. If that difference was below a certain threshold, an \( f_o \) data point was generated by the CUSTOM algorithm (NaN otherwise). The threshold was arbitrarily defined as 5% of an octave. Preliminary tests with a more rigorous threshold of 1/120 octave (ie, 10 musical cents), which approximates the just-noticeable difference for pitch perception in humans, considered decreased the usefulness of the CUSTOM algorithm, due to the great number of rejected data points even at slight levels of EGG signal quality degeneration.

Testing procedure
Including the CUSTOM algorithm, 14 algorithms were tested on the 147 EGG signals described earlier, resulting in a total number of 2058 observations. Prior to \( f_o \) calculation and GCI detection, the EGG signals were band-pass filtered twice using a third-order Butterworth filter with cutoff frequencies at 20Hz and 4800Hz. The second consecutive application of the filter was performed on the time-inverted input signal to negate phase distortion effects. The application of the band-pass filter was deemed appropriate, because comparable preprocessing steps would be performed in “real” data analysis situations. The cutoff frequencies were chosen carefully so as not to distort the analyzed signals.

Evaluation of performance
The output of \( f_o \) estimators and GCI detectors is fundamentally different in nature. Although the \( f_o \) estimators produce equidistantly spaced data points (every 1 ms in the case of this study) representing the time-varying (quasi-instantaneous) \( f_o \) information, the GCI detectors provide estimates of the time offsets of presumed glottal closure instants. To prevent adding any bias to the analysis (neither in favor of either \( f_o \) estimation nor GCI detection methods), we initially decided to compare the performance of all tested algorithms in both domains.

When comparing two frequencies, their difference in hertz is meaningless as an absolute value. A relative measure needs to be established instead. For the purpose of this study, the frequency differences between known and estimated \( f_o \) values were expressed in octaves:

\[
\Delta\text{oct} = \log_2\left(\frac{f_{\text{SYNTH}}}{f_{\text{EST}}}\right)
\]

For performance evaluation in the \( f_o \) domain, the glottal cycle information from the synthesized signal was converted to a time-series of \( f_o \) data at intervals of 1 ms. Based on this information, the following three parameters were calculated:

- A success metric, expressing the number of produced \( f_o \) data points in percent:
  \[
  \rho_{f_o} = 100 \frac{m}{n}
  \]

where \( n \) is the total number of possible data points (2000 for 2 seconds of synthesized sound) and \( m \) is the number of actually detected data points.

- Applying Equation 5, the average of the absolute differences between known \( f_o \) information from the synthesized signals and estimated \( f_o \) data was computed as follows:
  \[
  \mu_{f_o} = \frac{1}{n} \sum_{0}^{n-1} |\Delta\text{oct}[i]|
  \]

- Similarly, the standard deviation of \( f_o \) estimation was computed as:
  \[
  \sigma_{f_o} = \sqrt{\frac{1}{n} \sum_{0}^{n-1} (\Delta\text{oct}[i])^2}
  \]

The performance metrics parameters \( \rho_{GCI} \), \( \mu_{GCI} \), and \( \sigma_{GCI} \) for GCI detection were calculated in analogy to those for \( f_o \) estimation, with the difference that \( n \) was defined as the total number of glottal cycles in the respective synthesized signal.

Preliminary inspection of the algorithm performance data revealed no remarkable differences between the \( f_o \)-based \( (\rho_{f_o}, \mu_{f_o}, \text{ and } \sigma_{f_o}) \) and the respective GCI-based values \( (\rho_{GCI}, \mu_{GCI} \text{ and } \sigma_{GCI}) \), suggesting that conversion between \( f_o \) and GCI information did not introduce noteworthy artifacts into the data. Furthermore, there were no substantial differences of trends between the \( \mu_{f_o} \) and \( \sigma_{f_o} \) parameters. For these reasons, the remainder of this text focuses on the \( f_o \)-related parameters \( \rho_{f_o} \) and \( \sigma_{f_o} \) alone.

RESULTS
Detailed results of \( f_o \) detection from one representative signal are shown in Figure 2. An overview of the parameters \( \rho_{f_o} \) and \( \sigma_{f_o} \) for all analyzed scenarios is given in Figures 3 and 4. Detailed \( \mu_{f_o} \) success rates and \( \sigma_{f_o} \) scores for all analysis scenarios are provided in supplementary Tables S1 and S2.

With a few exceptions of extreme EGG signal modifications in the “compound” scenario and for extreme SNR values, most algorithms produced data for more than 90% of the possible 2000 data points per synthesized signal (Figure 3). Exceptions to this trend were found in the RAPT and DECOM algorithms, which typically had \( \rho_{f_o} \) values of about 90% and 80%, respectively. The CUSTOM algorithm deviated from its typical 95% \( f_o \) detection success rate when the random \( f_o \) variation \( \alpha \) was increased above 0.1 and when the amplitude modulation \( \beta \) was greater than 0.12, suggesting that above these critical values, the \( f_o \) readings from the two algorithms upon which the CUSTOM algorithm is based (ie, Praat “to Pitch (AC)” and SIGMA—see Methods) deviated by more than 5% of an octave.

Three of the analyzed algorithms (DYPSA, REAPER, and YAGA), all designed with the purpose of analyzing human speech, had problems recognizing \( f_o \) above
ca. 1000 Hz. Consequently, they were the worst-performing algorithms analyzed. The error benchmark $\sigma_{fo}$ for the DECOM algorithm was typically around 10% of an octave, rising considerably with increased random $fo$ variation $\alpha$. All other algorithms started out with acceptable $\sigma_{fo}$ ratings for EGG signals at lesser degrees of EGG signal quality distortion. However, increased random $fo$ variation had a tendency to gradually increase $\sigma_{fo}$ in all algorithms except DYPSA, REAPER, and YAGA. Overall, the CUSTOM and SIGMA algorithms had the best performance when testing for random $fo$ variation (Figure 4A).
For most of the algorithms, the occurrence of subharmonics appeared to be a crucial factor that led to abrupt increases in $\sigma_{f_o}$ over an amplitude modulation range of $0.1 > \beta > 0.24$ (Figure 4B). In each of these cases, the respective algorithm started to latch onto the subharmonic energy components in the signal. The respective threshold values were found at NDF: $\beta = 0.21$; Praat (AC): $\beta = 0.14$; Praat (SHS): $\beta = 0.24$; Praat (periodic cc): $\beta = 0.14$; Praat (periodic peaks): $\beta = 0.14$; RAPT: $\beta = 0.15$; and SWIPE: $\beta = 0.12$. As with random $f_o$ variation, the CUSTOM and SIGMA algorithms had the best performance with increased amplitude modulation.

No noteworthy trends were found with variation in amplitude drift, mains hum, or baseline drift—Figures 4C-E. Preliminary experiments conducted without band-pass filtering the signals before analysis revealed the same trends, even for baseline drifts.) The only exception was the NDF algorithm, which suffered an abrupt decrease in performance for amplitude drifts $\gamma < 0.42$, and the DECOM algorithm, which achieved reduced $\sigma_{f_o}$ values for $\gamma < 0.45$ and $\delta < 0.04$.

**FIGURE 3.** $f_o$ Data point resolution metric $\rho_{fo}$ for all analyzed algorithms and all synthesized EGG signals. (A)–(F) $\rho_{fo}$ as a function of the six simulated influence factors on EGG signal quality. (G) Effect of simultaneous change of all six influence factors.
Finally, typical EGG equipment noise seemed to be an important factor, influencing a number of algorithms (Figure 4F): There was an almost linear correlation between SNR of noise and \( \sigma_{\text{f}_0} \) in the “Praat (periodic peaks)” algorithm. More abrupt degenerations of performance (measured by increasing \( \sigma_{\text{f}_0} \)) were found for the following algorithms at respective thresholds: Praat (SHS): SNR = −5 dB; Praat (zeroes): SNR = 11 dB; RAPT: SNR = 1 dB; SIGMA: SNR = −1 dB; SWIPE: SNR = −3 dB. The CUSTOM, NDF, and Praat (AC) algorithms appeared to perform particularly well under the influence of noise, with terminal values of \( \sigma_{\text{f}_0} = 0.01 \) at an SNR of −5 dB.

Algorithm performance for linear combinations of the six influence factors described above are shown in the “compound” scenario illustrated in Figure 4G. The CUSTOM algorithm had a notably better performance (ie, lower \( \sigma_{\text{f}_0} \) values) than all other algorithms, particularly at higher degrees of EGG signal deterioration. This performance

**FIGURE 4.** \( \text{f}_0 \) Detection performance metric \( \sigma_{\text{f}_0} \), for all analyzed algorithms and all synthesized EGG signals. (A)–(F) \( \sigma_{\text{f}_0} \) as a function of the six simulated influence factors on EGG signal quality. (G) Effect of simultaneous change of all six influence factors.
success was, however, counterbalanced by the algorithm's lowered success rates $\mu_f$, (Figure 3G).

**DISCUSSION**

This study examines the performance of a number of $f_o$ and GCI detection algorithms when analyzing a special class of signals, that is, EGG signals with increasing complexity and at various stages of signal quality degradation. A total of six influence factors were assessed in this study: two inherent to the voice signal itself (random $f_o$, variation and subharmonics), and four types of signal degradations (amplitude and baseline drifts, mains hum, and typical EGG equipment noise). Mains hum, amplitude drift, and baseline drift all appeared to have a lesser influence on algorithm performance. The opposite was true for the other three factors—alteration in cycle-to-cycle variation (introduced by random $f_o$, variation), subharmonics (introduced by amplitude modulation of odd cycles), and typical EGG equipment noise all had a clear impact on algorithm performance. The somewhat disquieting main finding of this study is that there does not seem to exist one single “best” algorithm for analyzing EGG signals at various stages of complexity and degradation. For high-quality, low-noise EGG signals (eg, those typically acquired in excised larynx settings) the SIGMA algorithm seems to be the best choice. In signals with low SNRs, such as those collected in vivo from humans with a certain degree of fat tissue or phonating with incomplete glottal closure,24 or signals with suboptimal EGG electrode placement, the SIGMA algorithm does not appear to be the best choice. In those cases, NDF or Praat's AC algorithm would appear to be better suited. However, the performance of both NDF and Praat's AC algorithm is negatively affected by the occurrence of subharmonics.

In an attempt to consolidate these trends, a CUSTOM approach was introduced in this paper, combining the virtues of both SIGMA and Praat's AC algorithm. This CUSTOM algorithm showed the best performance overall (particularly in the “compound” scenario), but the improved performance came at the expense of discarding a large proportion of the analyzed data in situations where the outputs of the SIGMA and AC algorithms did not converge. This obvious trade-off between data quality and quantity can, to a certain degree, be controlled via the CUSTOM algorithm’s threshold setting (see Methods).

Some of the analyzed algorithms are intended to operate on certain types of signals.46 This may partially explain why the DYPSA, REAPER, and YAGA algorithms failed to produce meaningful data outside the typical $f_o$ ranges of human speech. Therefore, inferior performance of an algorithm in this study does not constitute a reason to conclude that the respective algorithm is inferior per se.

When studying the literature, the following pattern emerged: Most of the proposed $f_o$ and GCI detection algorithms were introduced by comparing their results with those from some other algorithms (differing across the various studies, and typically basing the tests on different input signals across different studies). Interestingly, in all of these cases, the respectively proposed algorithm had comparable or better performance than all other algorithms. Five nonmutually exclusive explanations can be found for this phenomenon:

(a) Owing to progress in the field of engineering the newly introduced algorithms become increasingly better over the years;
(b) Some algorithms work better for a certain type of data (eg, noisy data46 or special voice production types47,48) than others;
(c) Different methods of estimating algorithm performance result in different outcomes49;
(d) The authors of studies might have had a certain (potentially unconscious) a priori bias toward their “own” algorithm, which may have influenced them in choosing test data and competing algorithms for their performance tests; or finally,
(e) The authors may have made the mistake to train their algorithm on the chosen test data, leading to an overspecialized algorithm performance which cannot be generalized to other data sets.

Surprisingly, even studies that are only concerned with comparing algorithm performance (without introducing a new algorithm) do not converge to identical recommendations.48−54 suggesting that estimating algorithm performance might be as complex a task as $f_o$ or GCI detection itself. One way to address this issue is by consensually establishing databases of test signals with known properties. Advancing this notion, we have made all synthesized EGG signals utilized in this study available as supplementary materials.

Some of the considerations concerning standardizing algorithm performance evaluation also apply to this study. The $f_o$ range for synthesized signals was somewhat arbitrarily chosen to be in the range of 100−2000 Hz, in consideration of the human singing voice and the vocalizations of some nonhuman mammals. Furthermore, whereas three of the parameters for determining the synthesized EGG signals were chosen in relation to known value ranges (random cycle-to-cycle variation $\alpha$, amplitude modulation $\beta$, and SNR), the value ranges of the other three parameters had to be defined in an arbitrary fashion, based on the first author's long-term experience with EGG signals. Different values may naturally lead to different performance evaluation results. This is particularly true for the “compound” case, where all six parameters were varied in unison. In fact, preliminary tests with different, more extreme value ranges produced slightly different trends. For this reason, we have refrained from computing an overall metric of success across all synthesized signals. Such a metric would only apply to the given test data set and could not be generalized.
CONCLUSION
This study corroborates the insight that \( f_o \) detection is highly nontrivial.\textsuperscript{25,32} No single “best algorithm” was found for the special class of signals analyzed in this study. Thus, no recommendation for one single all-purpose \( f_o \) detection algorithm can be given. Rather, the nature of EGG data needs to be studied carefully before choosing an appropriate algorithm, and the insights from this study can help with that choice. Such an informed approach is recommended, rather than defaulting to a commonly used algorithm.

In summary, some main insights from this study are as follows: The researcher should never blindly trust a chosen \( f_o \) detection algorithm. Ex post facto, computed \( f_o \) data should always be assessed “by eye,” for example, via \( f_o \) traces superimposed upon narrow-band spectrograms. Furthermore, \( f_o \) data reported in the literature should not be taken at face value, particularly if the authors did not disclose (1) which \( f_o \) detection algorithm was chosen; (2) how the utilized \( f_o \) detection algorithm was chosen; and (3) whether (and how) the computed data were double-checked manually. There is an inherent degree of uncertainty and error in such data, due to the difficulties in automated \( f_o \) detection described in this paper.

Acknowledgments
This work is supported by an APART grant from the Austrian Academy of Sciences (to C.T.H.).

SUPPLEMENTARY DATA
Supplementary data related to this article can be found online at doi:10.1016/j.jvoice.2018.01.003.

REFERENCES