Inverse Filtering of Nasalized Vowels Using Synthesized Speech

*Christer Gobl and †James Mahshie, *Dublin, Ireland, †Washington, District of Columbia

Summary: The present study examines the extent to which increased nasal coupling affects estimates of glottal parameters derived from inverse filtering based on an all-pole assumption of the vocal tract. A series of steady-state tokens for five Swedish vowels were synthesized using the HLsyn quasi-articulatory synthesizer (Sensimetrics, Malden, MA). For each vowel, the parameter controlling the cross-sectional area of the nasal aperture, ***an***, was systematically varied, while the other HLsyn parameters were kept constant. The resultant pressure signal for each utterance was subsequently inverse filtered, and estimates were made of five glottal source parameters (**EE**, **RG**, **RA**, **RK**, and **OQ**) derived from fitting the Liljencrants and Fant source model to the inverse filtered signal. The results show that when analyzing nasalized vowels using inverse filtering based on an all-pole assumption of the vocal tract, the **RA** parameter estimate—a main determinant of the source spectral slope—can be adversely affected by nasal coupling. The errors in our estimates were particularly high for the high vowels: this was true not only for **RA**, but for all the parameters measured. However, with the exception of the distortion in the **RA** estimate, the effects were relatively small, regardless of the degree of nasal coupling.

Key Words: Nasalized vowels–Glottal parameters–Inverse filtering–Voice source.

INTRODUCTION

Inverse filtering is a technique for obtaining a glottal source signal by eliminating the acoustic effects of vocal tract resonances. It has been used to examine voice parameters of both normal[1–12] and disordered speech,[13–16] as well as in singing.[17–19] This filtering schema involves passing the speech signal through a filter whose transfer function is the inverse of the vocal tract transfer function. The resultant signal provides information about the voice source before being filtered by the supralaryngeal vocal tract.

Acoustically, oral vowels are the result of vocal tract resonances involving the configuration of the pharynx and oral cavity. Because of the single-tube-like characteristics of the vocal tract involved in producing oral vowels, these resonances can be represented by complex-conjugate pole pairs (each pair of poles corresponding to a formant in the speech signal). The inverse filtering of an oral vowel thus involves introducing complex-conjugate zeros, which effectively cancel the poles associated with the vocal tract resonances.

Nasalized vowels are the result of a different acoustic system, one involving not only of the pharyngeal and oral cavities but also of the nasal cavity. The nasal cavity is coupled to the oropharyngeal tract through the velopharyngeal port, which is formed by the lowering of the velum, thereby introducing a side branch to the vocal tract. The main acoustic consequence of such an acoustic arrangement is the creation of additional resonances and antiresonances (pole-zero pairs). Another effect is a shift in the frequencies of the resonances associated with the corresponding non-nasalized vowel (ie, the formant frequencies).

The specific frequencies of the additional poles and zeros associated with a nasalized vowel can vary considerably, depending in rather complex ways on the size of the velopharyngeal opening as well as on the particular shape of the oral and nasal cavities (for detailed accounts, see Refs 20,21). Acoustic coupling to the sinus cavities and asymmetries between the left and right nasal pathways may also play an important role.[22,23] Based on acoustic modeling, it can be shown that the frequency of the first pole in the transfer function of the nasalized vowel will always be lower than the frequency of the first zero.[20(p152)] Furthermore, when the coupling is small, the pole and the zero of the additional pole-zero pairs will be in close proximity, and consequently their effect on the transfer function is small, as the poles and zeros partially cancel each other. However, an increase in the velopharyngeal opening causes a widening of the frequency gap of the pole-zero pairs, resulting in a greater change in the transfer function.

An important acoustic cue for the percept of nasalization seems to be a reduced prominence of the spectral peak in the region of the first formant frequency.[20,24–26] This reduction is thought to be due to a combination of factors such as the greater damping of resonances in the nasal cavity as well as the canceling effect of the first antiresonance when coming into close proximity with $F1$.[20] Alternatively, one can view the characteristics of the nasalized vowel in the frequency region of $F1$ to be the result of a pole-zero-pole constellation made up of $F1$ and the first pole-zero pair.[27,28]

The all-pole assumption typically made about the vocal tract in the inverse filtering schema is in theory limited to utterances that are non-nasalized. There have been attempts to inverse filter nasalized utterances using a more complex filter schema.[29] However, this approach is difficult, particularly when a formant and a nasal zero fall in close proximity because they will partially cancel each other. Furthermore, care must be taken to

Accepted for publication September 13, 2012.

C.G. and J.M. contributed equally to this study.

This research was supported by the Science Foundation Ireland (Grant 09/IN.1/I2631) as part of the FASTNET project.

From the *Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Dublin, Ireland; and the †Department of Speech and Hearing Science, The George Washington University, Washington, District of Columbia.

Address correspondence and reprint requests to Christer Gobl, Centre for Language and Communications Studies, Arts Building, Trinity College Dublin, College Green, Dublin 2, Ireland. E-mail: cegobl@tcd.ie

Journal of Voice, Vol. 27, No. 2, pp. 155-169

0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2012.09.004

avoid canceling an additional resonance without canceling the associated antiresonance, as otherwise the inverse filtering would overcompensate for the effect of the additional resonance.

Others (for example, Ananthapadmanabha)[30] have suggested that the all-pole model provides fairly robust results for utterances that are moderately nasalized. No research to date has been reported that quantifies the degree and nature of changes in glottal source parameter estimates from inverse filtered utterances containing varying degrees of nasalization. Accordingly, the present study examines the extent to which systematically varying the degree of nasal coupling affects the estimates of glottal parameters when derived from inverse filtering based on an all-pole approximation of the vocal tract transfer function.

It is clear from the above description that the spectrum of nasalized vowels is rather complex, and this makes inverse filtering of such sounds considerably more complicated than the inverse filtering of oral vowels. This can pose problems in the use of inverse filtering to examine voice characteristics of some disordered speach. For example, Thompson and Murdoch[31] report that the speech of patients with upper motor neuron lesions following cerebrovascular accidents often contains some degree of nasal coupling, while others suggest that voice problems are typical of this population.[32,33] Theodoros et al[34] studied a group of 20 individuals with post-traumatic brain injury and identified hypernasality in 95% of the participants, with most in this group (63%) demonstrating moderate-to-severe disturbances. Additional evidence examining dysarthria in patients with multiple sclerosis also reported a high incidence of phonatory and nasal coupling disruptions.[35]

Similarly, the speech of deaf individuals is often described as having both voice and nasal quality disruptions.[36,37] Mahshie and Öster[16] noted this in their study of glottal volume velocity obtained through inverse filtering and limited their analysis to utterances that were produced with no appreciable nasal airflow. However, this poses potential issues regarding the representativeness of speech samples because only utterances with minimal nasal coupling can be analyzed. There is thus a challenge of obtaining information about a range of disordered speech through inverse filtering in populations for which nasality may be an associated disruption.

Similar challenges may arise in the analysis of the singing voice. Although some studies have pointed to a lesser degree of vowel nasalization in classical singing,[38,39] other studies suggest that the velopharyngeal port is indeed actively adjusted in vowel production to control the timbre.[40,41]

Inverse filtering and voice source analysis

Inverse filtering can be applied to either the volume velocity or the pressure wave. In examining the volume velocity waveform, investigators have identified a number of parameters related to voice quality such as pulse amplitude, closed quotient, speed quotient, and maximum flow declination rate.[6,13,42]

Alternatively one can inverse filter the pressure wave. The primary difference in working with the pressure wave is the need to consider the radiation function. The effect of sound radiation at the lip opening can be approximated by a first-order differentiation of the oral volume velocity.[20] Consequently, integration of the signal would be required to cancel the filtering effect of the radiation function. However, some choose not to compensate for the radiation function: the signal thus resulting from the inverse filtering is referred to as the differentiated glottal flow (or the glottal flow derivative). In terms of the slope of the spectrum, higher frequencies are emphasized by 6 dB per octave in the differentiated glottal flow compared with the true glottal flow.

There are advantages and limitations to using either of these signals for examining the voice source. The flow signal permits measurement of absolute values of airflow rate, something that cannot be obtained from the pressure signal. The flow signal, however, has a limited bandwidth, thus providing limited information about the spectral attributes of the source. Additionally, an airflow measuring system requires a mask, and transducers to obtain the flow signal.

A significant reason for obtaining the voice source waveform (either from the oral airflow or the pressure wave) is to gain insight into either the physiological underpinnings of these signals or the acoustic characteristics that result. Although qualitative examination of the waveforms can yield general descriptions (such as breathy or efficient); more precise, quantitative descriptions are desirable by providing numerical interpretations to variations in waveform characteristics. These can be made either directly on the source signal[6,43] or by matching a source model to the output of the inverse filter, and then deriving parameters from the modeled waveform.[44] An advantage of this latter approach is that it facilitates parameter optimization that takes into account both time and frequency domain information. This matching procedure is discussed in greater detail below.

Although numerous voice source models have been proposed, in this work we use the Liljencrants and Fant (LF) model,[45] which has gained popularity for its ability to accurately capture the salient features of the glottal waveform with a minimal number of control parameters. The LF model is defined in terms of differentiated glottal flow and an example showing two LF model pulses is displayed in the bottom panel of Figure 1. The top panel of the figure shows the corresponding glottal flow. The same figure also shows the glottal parameter definitions used in the current analysis. For detailed descriptions of this model, see Refs.[44,46–48]

Once the LF model has been matched to the pulses of the glottal waveform, various parameters can be estimated from the time and amplitude information of the model. Apart from the fundamental frequency, the LF model pulse is determined by four parameters. Thus, in addition to f_0, we normally measure at least four parameters from the matched LF pulses: the parameters we would typically measure are **EE**, **RA**, **RG**, and **RK**. Although these parameters are general glottal source parameters with definitions that do not depend on any particular source model, they are often referred to as the "LF parameters," despite the fact that they are not the actual parameters of the functions defining the model.

In this study, we analyzed five parameters: **EE**, **RA**, **RG**, **RK**, and **OQ** (for details, see Methods section). It is worth

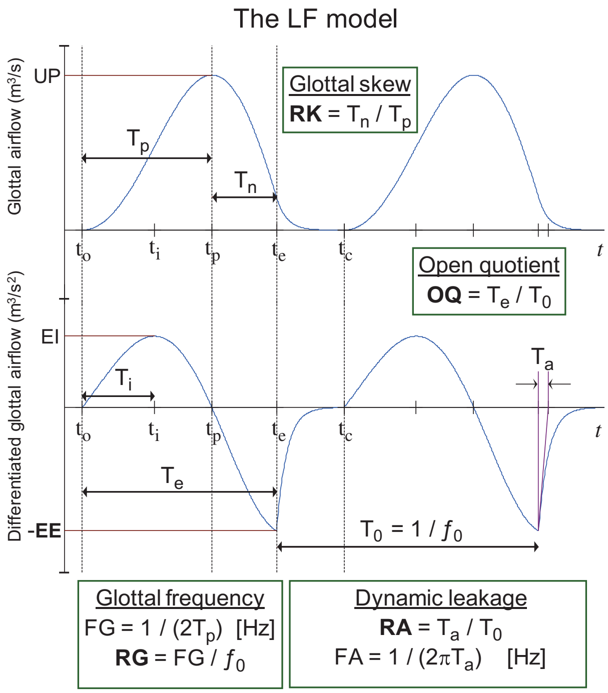

FIGURE 1. Glottal source parameter definitions. The lower waveform shows two LF model pulses of differentiated glottal flow. The upper waveform shows the corresponding glottal flow pulses.

pointing out that the **OQ** value can be derived from the values of **RG** and **RK** and is therefore redundant. However, as the **OQ** parameter is very commonly used, it was deemed beneficial to include data also for this parameter.

For ease of reference, this set of parameters shall be referred to as the LF parameters, as it helps to keep them distinct from the synthesis parameters used in HLsyn (Sensimetrics, Malden, MA) and KLSYN88, which are discussed in the following sections. Note also that the LF parameters are shown in bold upper case, in contrast to the HLsyn parameters which are shown as bold italic lower case (eg, ***an*** and ***f1***) and the KLSYN88 (including the KLGLOTT88) parameters, which are shown as plain upper case (eg, TL and FNP).

Synthetic speech and analysis by synthesis

Studying the effectiveness of inverse filtering with speech that is nasalized is challenging. The use of natural speech does not enable researchers to know exactly the characteristics of the glottal source signal. Moreover, there are limited approaches to controlling the degree of nasal coupling to examine the impact of varying degrees of nasal coupling on the filtered output. The use of synthetic speech provides an alternative that addresses both these concerns. It is possible to carefully control the glottal source used in the synthesis (which is known *a priori*), and it is possible to systematically vary the degree of nasal coupling.

The HLsyn synthesizer[49,50] uses a model based on articulatory and acoustic parameters, such as areas of constrictions of the vocal tract, change in oral cavity volume, tissue compliance, subglottal pressure, and natural frequencies of the tract. These parameters generate the control parameters of the Klatt formant synthesizer, KLSYN88,[51] through a set of mapping relations and a low-frequency model of the aerodynamics of articulation. HLsyn has been used in a variety of studies, including the investigation of individual speaker characteristics,[52] production and perception of British English,[53] the speech characteristics of deaf individuals,[54] and the naturalness of an American English text-to-speech system.[55] An appealing aspect of HLsyn is that "the model includes constraints on source-filter relations that occur naturally during speech production."[50]

The HLsyn controls the Klatt synthesizer through 13 higher-level parameters shown in Table 1. In addition, there are a series of user-selectable speaker constants that represent the default values of the constants used within the equations that convert HLsyn parameters into KLSYN88 parameters. The speaker constants are related to a number of factors such as the physical characteristics of the individual, the neutral vowel formant frequencies and bandwidths, constants related to the source characteristics, and scale factors used during the calculation of the KLSYN88 parameters.[50] Each higher-level HLsyn parameter represents either an acoustic parameter that is related directly to articulation (eg, ***f1***, ***f2***, and ***f3***) or an articulatory attribute of the vocal tract (eg, ***an*** and ***al***, the cross-sectional aperture

TABLE 1.
Description of HLsyn Parameters

f1–f4	First four natural frequencies of vocal tract, assuming no narrow local constrictions (Hz)
f0	Fundamental frequency due to active adjustments of vocal folds (Hz)
ag	Average area of glottal opening between the membranous portion of the vocal folds (mm^2)
ap	Area of the posterior glottal opening (mm^2)
ps	Subglottal pressure (cm H_20)
al	Cross-sectional area of constriction at the lips (mm^2)
ab	Cross-sectional area of tongue-blade constriction (mm^2)
an	Cross-sectional area of velopharyngeal port (mm^2)
ue	Rate of increase of vocal tract volume (cm^3/s)
dc	Change in vocal fold or wall compliances (%)

Adapted from Hanson and Stevens.[50]

size at the velum or lips, respectively). Additional parameters control elements of the glottal source as well (eg, ***ag*** and ***ap***). Thus, the synthesizer can be used to carefully control elements of articulation and phonation.

In the present study, HLsyn was used to generate a series of utterances with a consistent glottal source function and with varying degrees of nasal aperture size. The aim was to examine the effect of varying degrees of nasal coupling on estimates of the LF glottal source parameters. The inverse filtering schema did not make any accommodation for nasal coupling, that is, the filtering assumed an all-pole model of the vocal tract. Thus, the present work examines the impact of nasal coupling on inverse filtering when there are violations of the assumptions underlying this approach to examining voice parameters.

METHODS

Synthesis of utterances

The HLsyn synthesizer[49,50] was used to generate the speech utterances in which the degree of nasal coupling was systematically varied. As noted previously, each high-level parameter of HLsyn controls multiple KLSYN88 parameters.

The degree of nasal coupling is set by a parameter called ***an***, which varies the size of the nasal aperture from 0 to 100 mm^2. The ***an*** parameter causes changes primarily in three KLSYN88 parameters: the frequency of the nasal pole (FNP), the bandwidth of the nasal pole (BNP), and the frequency of the nasal zero.

To examine the effect of different degrees of nasal coupling on different vowels, the following vowels were synthesized: /a/, /i/, /u/, /ɑ/, and /ə/. The formant frequencies, given in Table 2, are based on Swedish vowels and were used as input values for the HLsyn parameters ***f1***, ***f2***, ***f3***, and ***f4*** (Table 2). Apart from the nasal aperture (***an***), all other HLsyn parameters were determined by the default settings as specified by the speaker constants for a male voice.[50(pp1179–1180)]

Although KLSYN88 offers a choice of three different glottal source models (one being a modified version of the LF model), only the default source model, KLGLOTT88, can be used with HLsyn. Hence, the synthesized utterances in the present study used the KLGLOTT88 model to generate the glottal source signal. This model has three control parameters: AV, amplitude of voicing, which sets the overall amplitude of the pulses; OQ, the open quotient; and TL, which controls the amount of tilt of the source spectrum. The TL value is the additional attenuation in decibels at 3 kHz.[51]

TABLE 2.
Formant Frequency Values Used in HLsyn to Synthesize the Utterances

Vowels	F_1 (Hz)	F_2 (Hz)	F_3 (Hz)	F_4 (Hz)
/a/	750	1250	2500	3350
/i/	300	2200	3150	3750
/u/	300	600	2350	3250
/ɑ/	600	950	2550	3330
/ə/	500	1500	2500	3500

Approximately 500 milliseconds of each utterance were synthesized and then analyzed as described in the following sections. To examine the full extent of the effect of nasal coupling on the source estimates, we used the full range of ***an*** values of the synthesizer, that is, from 0 to 100 mm^2. Each of the five vowels in Table 2 was synthesized using 15 different settings of ***an***, where the specific values used were 0, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, and 100 mm^2. Note that these ***an*** values were not based on any particular area measurements of the nasal aperture but were simply selected to cover the widest possible range of nasal coupling.

According to Stevens,[21(p18)] for the production of sounds involving the nasal cavity, the velopharyngeal opening is typically in the range of 20–80 mm^2. The range used in our synthesized vowels therefore amply covers the range likely to occur in various typical and atypical speakers. This range of nasal aperture values is consistent with those used by others[56,57] and reflects the full range of possible velopharyngeal adjustments.

Glottal waveform analysis

To examine the glottal waveform characteristics of each synthesized utterance, an interactive voice source analysis system was used, which was developed at Trinity College Dublin.[58] The system incorporates a series of programs that provide automatic and interactive analysis of source and vocal tract filter parameters. This method has been used in many previous analyses of the voice source, including not only vowels but also a wide range of voiced consonants.[4,5,8,11,59]

The first set of programs enable inverse filtering of the synthesized vowels to obtain an estimate of the differentiated glottal flow waveform. Initially, the formant frequencies and bandwidths are derived automatically. These values can subsequently be adjusted to optimize the accuracy of the inverse filtering using an interactive program. To illustrate the procedure, Figures 2 and 3 show the analysis for two /u/ vowels, one without nasalization (Figure 2, ***an*** $= 0$ mm^2) and one with maximum nasalization (Figure 3, ***an*** $= 100$ mm^2).

The inverse filtering strategy used in this study uses five antiformants, one for each of the formants used in the synthesis. Furthermore, formant values were constrained to remain realistic for a given vowel quality. For instance, the antiformant used for $F5$ could not be "reassigned" to cancel the nasal pole. Given these constraints, the optimization criterion in the time domain was to achieve maximum cancellation of oscillations in the glottal closed phase. In the frequency domain, the aim was to maximally cancel the spectral peaks, while retaining realistic spectral continuity. In Figure 3, the effect of inverse filtering the nasalized vowel using only zeros can be noted: the main distortion of the source spectrum is the spectral dip in the region of 950 Hz.

A different interactive program then permits adjusting the LF model to the glottal pulses obtained from the inverse filtering (Figures 2D and E, and Figures 3D and E). The matching process involves manipulating the shape of the LF model by adjusting five cursors in the time domain, specifying the time points

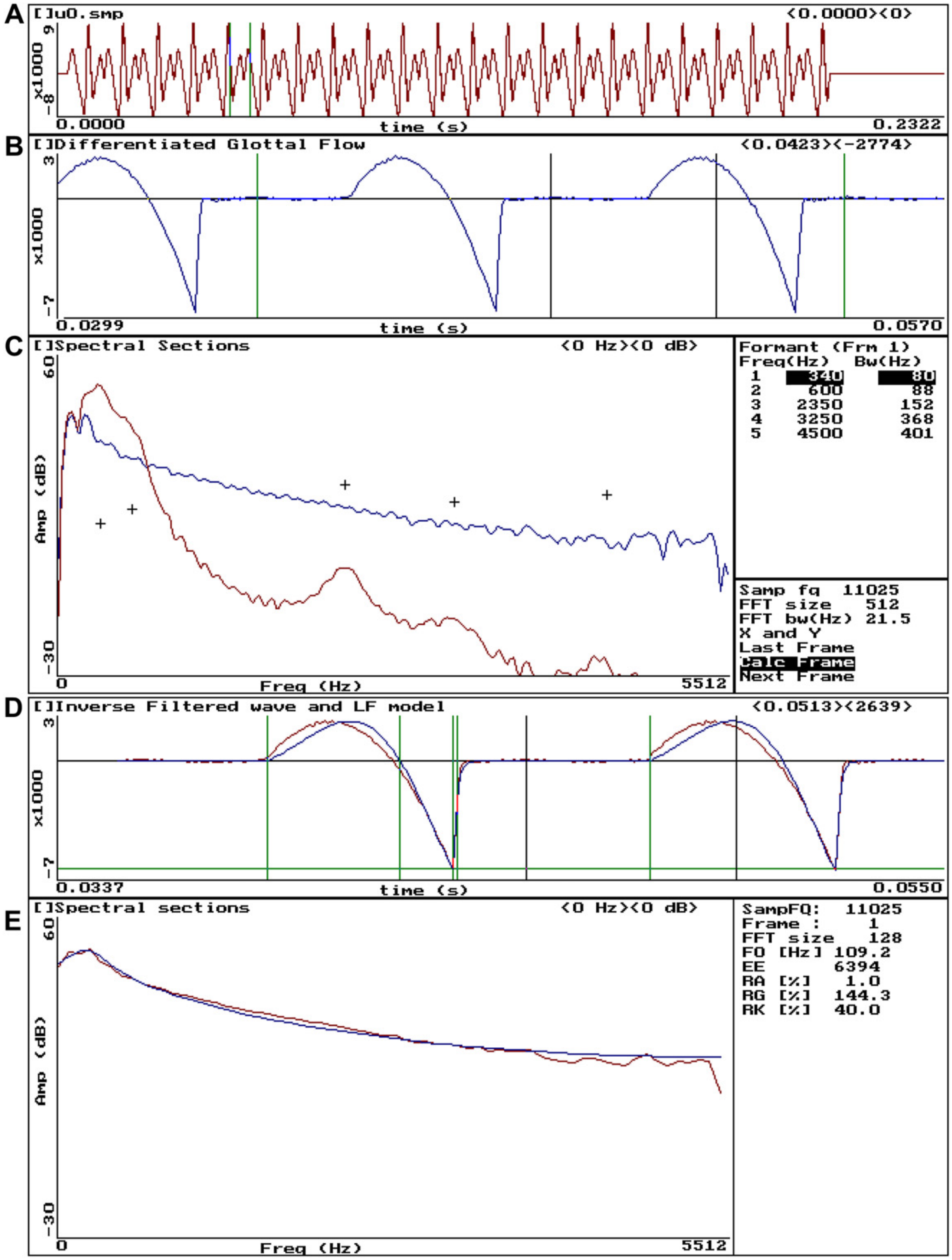

FIGURE 2. Screen display of the interactive inverse filtering software showing the analysis of the synthetic non-nasalized /u/ vowel (***an*** = 0). The screen display of the interactive inverse filtering program is presented in the upper half of the figure. **A.** The speech waveform of the whole utterance, where the vertical line indicates the cycle of the waveform being analyzed. **B.** The output signal of the inverse filter (ie, the differentiated glottal flow estimate). **C.** The amplitude spectra of the speech waveform (*in red*) and the differentiated glottal flow (*in blue*). The points (*plus signs*) in this window represent the formants, and their locations determine the complex zeros of the inverse filter. Each of these points can be moved horizontally or vertically to change the frequency and bandwidth, respectively. **D**. The LF model waveform (*in blue*) matched to the output signal of the inverse filtering (*in red*). **E**. The corresponding amplitude spectra of the waveforms in **D**.

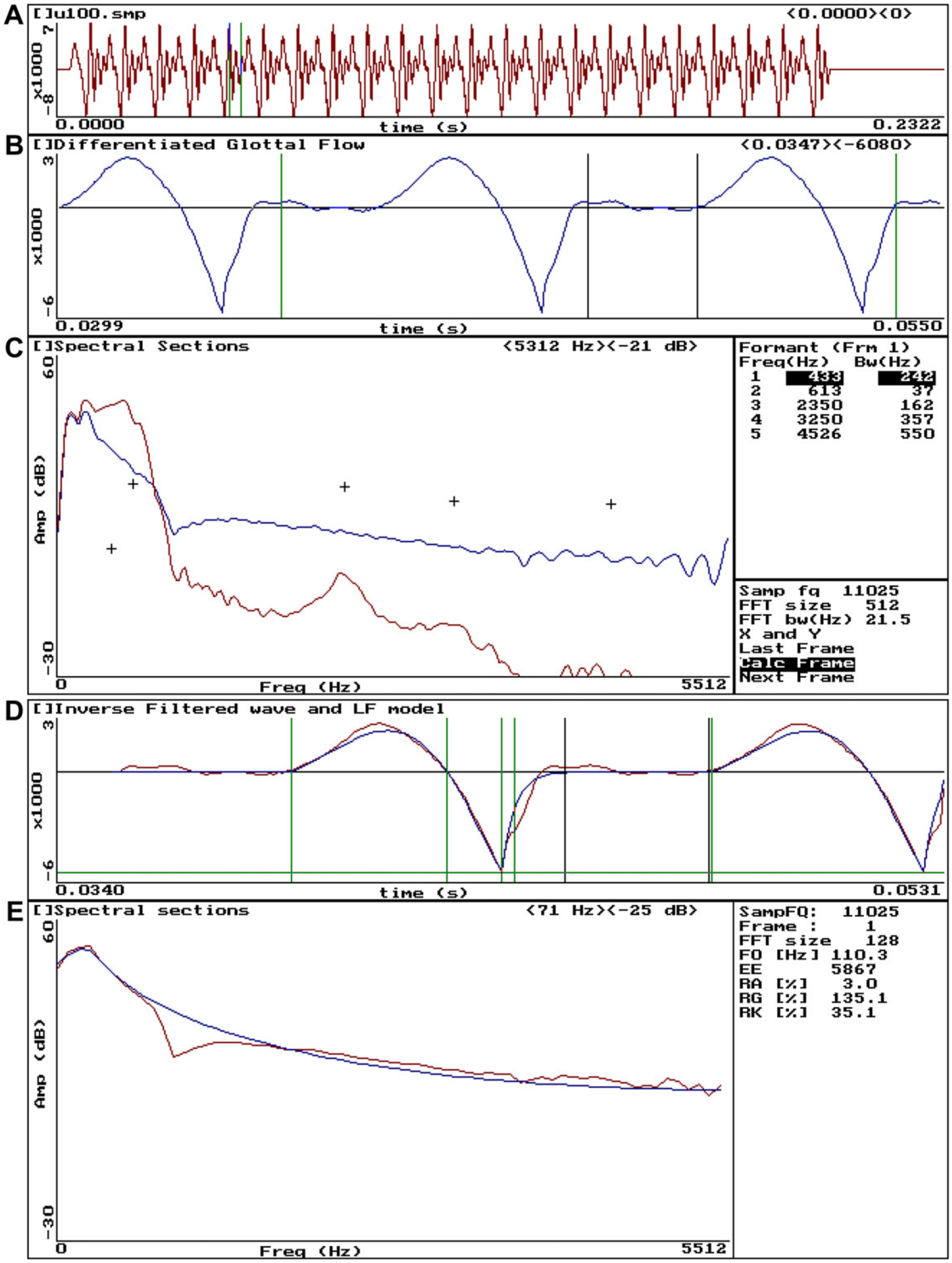

FIGURE 3. Screen display of the interactive inverse filtering software showing the analysis of the synthetic nasalized /u/ vowel (***an*** = 100 mm^2). The upper half of the figure shows the inverse filtering displays and the lower half shows the model matching displays (see Figure 2 caption for details).

t_o, t_p, t_e, and t_r (=t_e + TA), and the amplitude value **EE** (Figure 1). From these points, the LF model waveform (*in blue*) is generated and superimposed on the inverse filtered waveform (*in red*). It is important to note that the matching process does not simply involve model fitting in the time domain. An essential and integral part of the modeling is the direct comparison it allows for the amplitude spectrum of the model and that for the derived glottal pulse (Figures 2 and 3D and E). The frequency domain matching is particularly important for capturing the main acoustic characteristics of the source function. For example, achieving good estimates of certain source parameters, such as **RA**, can often be difficult without the spectral information, as the closest temporal fit does not necessarily provide the best spectral match.

The optimization process of the model matching was as follows. The t_e and **EE** values were always set as indicated by the glottal waveform. For our synthetic stimuli, this was unproblematic as there was always a clearly defined point of excitation occurring at the maximum negative amplitude of the pulse. Thus, nearly exact measures could be obtained (the only limitation being the screen resolution). These two values were not allowed to be further adjusted, say, to produce a visually closer model fit. However, the other three time points, t_o, t_p, and t_r, were allowed to be fine-tuned to optimize the match in the time and frequency domains. In the time domain, priority was given to capturing the shape of the waveform in the vicinity of the main excitation, focusing particularly on the slopes just before and after it. In the frequency domain, the match in the mid to upper end of the spectrum (broad levels and slope) and the level in the region of the glottal formant were given special attention. The accuracy of the glottal opening phase (between t_o and t_p) was given less priority as an exact fit in this region has relatively less impact on the source spectrum. It can be noted from Figure 3 that a reasonably good temporal and spectral match can be obtained for the nasalized vowel despite the distortion caused by the limitations in the inverse filtering.

Because each utterance was essentially a steady-state vowel, it was deemed sufficient to analyze only three cycles from each utterance and average these results. For each ***an*** condition, the resultant signals were subsequently analyzed to determine the following five parameters (for further details on these parameters, see Ref.[44] for a comparison with the set of glottal parameters typically used in *Journal of Voice*, see Appendix 2):

EE—excitation strength is a measure of the strength of the main glottal excitation as defined by the *negative* value of the amplitude at the time of maximum glottal pulse discontinuity (which means that **EE** is a positive value, cf. Figure 1). This amplitude is typically equivalent to the maximum negative amplitude of the differentiated glottal pulse. Changes in **EE** affect the overall amplitude of the source spectrum: an increase in **EE** boosts all source components equally, with the exception of the very lowest ones, particularly the first harmonics. The amplitudes of these lowest harmonics are more influenced by the pulse shape rather than by **EE**.

RA—dynamic leakage is a measure of the effective duration of the return phase, TA (Figure 1), normalized to the fundamental period T_0. The **RA** measure relates to the amount of "dynamic leakage," that is, the residual flow during the return phase, which occurs from the time of the main excitation to the time of complete or maximum closure. Differences in dynamic leakage are important acoustically because they affect the spectral slope of source. An increase in the dynamic leakage would result in a relative weakening of the higher components of the source spectrum.

RG—normalized glottal frequency is a measure of the glottal frequency, FG, normalized to the fundamental frequency, f_0, where FG is the characteristic frequency of the glottal pulse during the open phase.[60] **RG** affects the relative amplitudes of the very lowest components in the source spectrum. An increase in **RG** would result in attenuated levels of the lowest end of the source spectrum. The higher the **RG** value, the more it will approach the frequency of the second harmonic and the more it will contribute to strengthen it rather than the first harmonic.

RK—glottal pulse symmetry is a measure of the degree of symmetry of the glottal pulse. **RK** is defined as the duration of the closing portion of the pulse (from the time point of peak glottal flow to the time point of the main excitation) as a proportion of the duration of the opening portion of the pulse (from the time point of glottal opening to the time point of peak glottal flow). The more skewed (asymmetrical) the pulse, the lower the **RK** value. An increase in **RK** (ie, a more symmetrical pulse) results in greater spectral dips, that is, a less even spectral slope.[46] A more symmetrical pulse will also result in a strengthening of the spectral components in the region of the glottal frequency, FG, as a more symmetrical pulse effectively means that the glottal formant bandwidth is reduced.

OQ—open quotient is a measure of the proportion of the glottal cycle for which the glottis is open. The **OQ** values presented here, however, do not include the duration of the return phase, that is, **OQ** is defined as T_e/T_0 (Figure 1). This definition of the open quotient conforms to the one used in KLSYN88. As noted previously, the **OQ** measure is determined by **RG** and **RK** and can be derived as follows: $\mathbf{OQ} = (1 + \mathbf{RK})/(2\mathbf{RG})$. Thus, **OQ** tends to be positively correlated with **RK** (ie, **OQ** tends to increase with increasing **RK**) and negatively correlated with **RG** (ie, **OQ** tends to decrease with increasing **RG**). In the frequency domain, changes in **OQ** mainly affect the amplitude of the lower components. In particular, there tends to be a strong positive correlation between the **OQ** value and the amplitude of the first harmonic.[47]

Adjustment of glottal parameter estimates

Ideally, we would have TL and AV values constant across all ***an*** conditions. However, the actual values of the KLSYN88 parameters TL and AV used in the synthesis are derived from the settings of the HLsyn parameters and are thus not directly set by the user.

Because TL directly affects **RA** and AV directly affects **EE** (see the following sections), these differences would clearly introduce variation in the glottal signal that is not attributable to inverse filtering. As a result, it was necessary to adjust the glottal parameter estimates for some of the data. Adjusting the known alterations that the synthesizer made to these parameters acted to "normalize" them. These adjustments are thus "correcting" alterations made directly to these LF parameters by HLsyn that would be viewed as a distortion if not accounted for in our analysis. The rationale and approach taken for these adjustments is described in the following sections and also in Appendix 1.

The TL values used as input to the KLSYN88 for the different utterances ranged from 2 to 8 dB. As the TL value affects the source spectral slope, the differences in TL would obviously affect our **RA** estimates. Thus, all our **RA** values were normalized to correspond to TL = 5 dB. The way the normalized **RA** values were derived is described in Appendix 1.

As pointed out previously, there was also some variation in the AV values set by HLsyn. Consequently, our **EE** estimates needed to be adjusted, as any change in AV would result in a corresponding change in **EE**. For most utterances, AV = 60 dB, but in some cases, AV was reduced to 59 dB. In these cases, 1 dB was added to the **EE** value, so as to normalize our **EE** estimates to correspond to AV = 60 dB.

With regard to our estimates of **RG**, **RK**, and **OQ**, no adjustments were necessary, as the OQ parameter (of the KLGLOTT88 model) remained at 50% for all synthesized utterances.

RESULTS

The measures of the LF parameters presented here are based on the mean values from the analysis of three cycles of each vowel, varying across the cross-sectional areas of nasal aperture (***an***). Note that values of **RA**, **RG**, **RK**, and **OQ** are shown as percentages. Thus, they were calculated as follows: **RA** = $100(\mathrm{TA}/T_0)\%$, **RG** = $100(\mathrm{FG}/f_0)\%$, **RK** = $100(T_n/T_p)\%$, and **OQ** = $100(T_e/T_0)\%$.

Variation in EE

Figure 4 presents the results for the **EE** (excitation strength) parameter, after normalization. As mentioned previously, this normalization involved adding 1 dB to the original **EE** values from the analysis for the utterances where HLsyn imposed a lowering of AV by 1 dB. The variation in **EE** was, on the whole, small as a function of ***an***. The general trend was for a small reduction in the **EE** estimates with increasing nasal coupling. This reduction was larger for the high vowels (1.9 dB for /u/ and 1.0 dB for /i/), but most of the change occurred between 0 and 10 mm^2 nasal aperture. For the non-high vowels /a/, /ə/, and /ɑ/, the change was 0.6 dB or less. Note that for utterances with ***an*** = 0, there were differences among the five vowels of approximately 0.5 dB.

Variation in RA

As noted earlier, the **RA** parameter is a measure corresponding to the residual flow during the return phase, which occurs from the time of the main excitation to the time of complete or maximum closure. The data in Figure 5 show some striking effects on the **RA** estimates (here normalized to correspond to TL = 5, as described in the Methods section). As was seen for **EE**, the high vowels were most affected. For /i/, **RA** values ranged between 1.4% (***an*** = 0) and 6.5% (***an*** = 90 mm^2). For /u/, the trend was similar, but the **RA** increase plateaued between 3% and 4% beginning at ***an*** = 30 mm^2. For /a/ and /ə/, although differences were smaller, changes in relative terms were nevertheless noteworthy and would have considerable impact on the source spectrum. The smallest variation in **RA** was for /ɑ/. The **RA** estimates for utterances with no nasal coupling ranged between 1.0% and 1.4%.

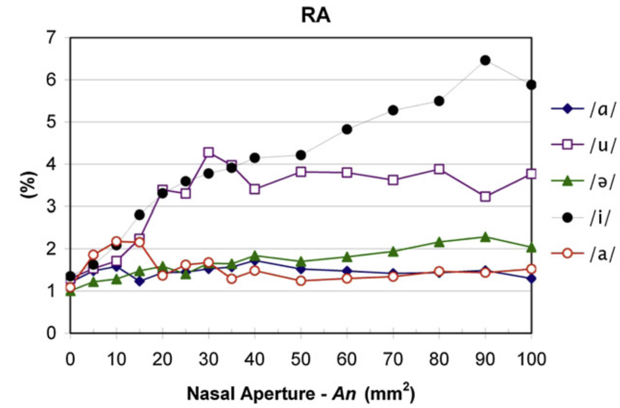

FIGURE 5. **RA** changes associated with varying cross-sectional area of nasal aperture (***an*** = 0–100 mm^2).

Variation in RG

As noted earlier, **RG** represents the normalized glottal frequency, FG/f_0, where the glottal frequency, FG, is the characteristic frequency of the glottal pulse during the open phase. Figure 6 shows that the variability of **RG** estimates for ***an*** = 0 were very small, ranging from 143% to 144%. As was seen with other parameter estimates, **RG** values tended to vary to differing degrees for different vowels as ***an*** increased. The **RG** variations were quite similar throughout the range of ***an*** values for the vowels /a/, /ə/, and /ɑ/, with variations tending to be within 2–3% of the ***an*** = 0 condition. For

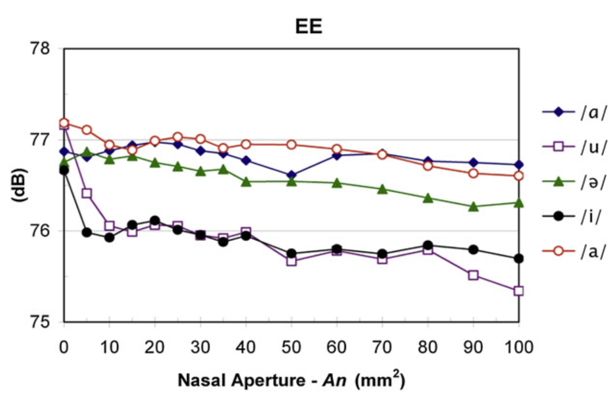

FIGURE 4. **EE** changes associated with varying cross-sectional area of nasal aperture (***an*** = 0–100 mm^2).

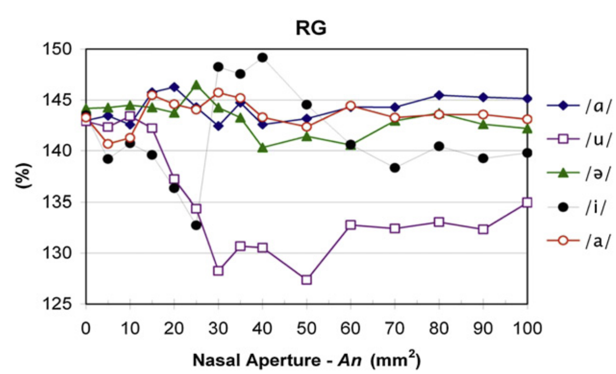

FIGURE 6. **RG** changes associated with varying cross-sectional area of nasal aperture (***an*** = 0–100 mm^2).

/u/, **RG** values decreased for ***an*** values between 15 and 30 mm^2. Beyond 30 mm^2, the parameter values were fairly constant, although lower than for the other vowels. For /i/ utterances, **RG** varied in a rather complex way. Until ***an*** reached 25 mm^2, RG values tended to drop. At ***an*** = 30 mm^2, the **RG** value then increased by approximately 10% and stayed relatively high until ***an*** = 40 mm^2. Then, **RG** values again dropped, returning to approximately the same **RG** level as recorded for the smallest ***an***.

Variation in RK

It was noted previously that **RK** is a parameter that measures the degree of pulse symmetry, that is, it relates inversely to pulse skewness—as skewness of the pulse increases, **RK** decreases. As shown in Figure 7, the variation in **RK** as a function of ***an*** was relatively small. For /a/, /ə/, and /ɑ/, the variation is minimal (less than 3%) both with increasing degrees of coupling and among different vowels. Again, the high vowels exhibited the most variation with increasing ***an***, there being a tendency for **RK** values to decrease (ie, to show increased glottal skewness) with increasing ***an***. Some of the change in **RK** is of course a direct consequence of the variation in **RG**. In the case of no nasal coupling, **RK** values were very consistent; values varied only between 39.3% and 40.2%.

Variation in OQ

As noted earlier, **OQ** represents the proportion of the glottal cycle when the glottis is open. As seen in Figure 8, all vowels showed fairly consistent estimates of **OQ** values for ***an*** values ranging from 0 to 15 mm^2. For the vowels /a/, /ɑ/, and /ə/, consistent estimates of **OQ** were maintained throughout the range of ***an*** values. The **OQ** for the vowel /u/ tended to increase between 15 and 30 mm^2, beyond which the parameter tended to be relatively constant, but higher than the **OQ** for the other vowels. The **OQ** for the vowel /i/ increased between 15 and 25 mm^2 and then dropped approximately 5% between 25 and 30 mm^2. In addition, **OQ** estimates for /i/ stayed relatively low between 30 and 50 mm^2 and then approximated the **OQ** estimates for /a/, /ə/, and /ɑ/ for ***an*** higher than 60 mm^2. Given the fact that the **OQ** measure is completely determined by the values of **RG** and **RK**, it is not surprising that the vowels /a/, /ɑ/, and /ə/ showed consistent estimates for **OQ** throughout the range of ***an*** values as this was also the case for the **RG** and **RK** estimates for these vowels. For /u/ and /i/, the variation in **OQ** appears mainly a consequence of the variation in **RG** (ie, variation in the duration of the opening portion of the glottal cycle). Also for **OQ**, the results were very consistent for utterances with no nasal coupling; these values were between 48.3% and 49.1%.

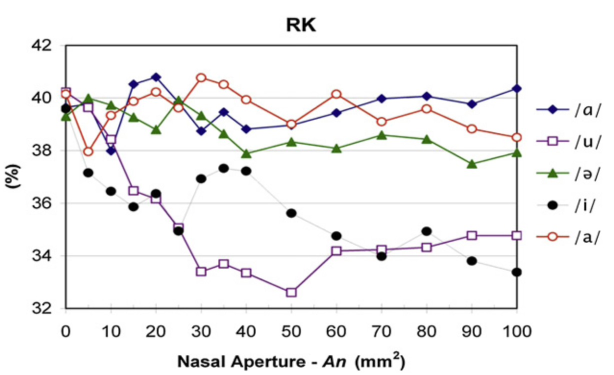

FIGURE 7. **RK** changes associated with varying cross-sectional area of nasal aperture (***an*** = 0–100 mm^2).

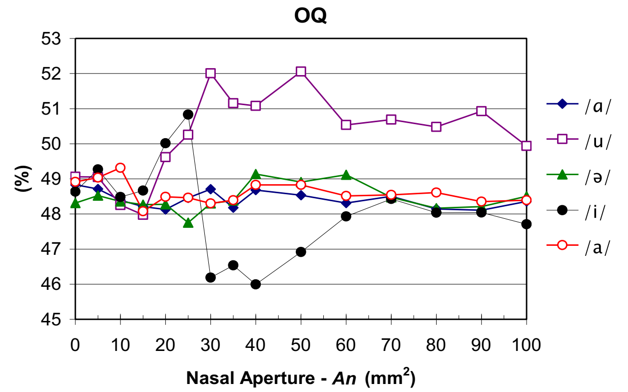

FIGURE 8. **OQ** changes associated with varying cross-sectional area of nasal aperture (***an*** = 0–100 mm^2).

Maximum errors in the voice source estimates

Perhaps the most notable finding is that the range of values for all the parameter estimates appears to be greater for the high vowels /i/ and /u/ than for the non-high vowels /a/, /ə/, and /ɑ/. Figure 9 shows the maximum errors found in the parameters over the whole range of ***an*** values. Maximum error of a parameter is defined as the maximum variation of that measure from the non-nasal utterance (***an*** = 0) for each vowel. For **EE** estimates, the maximum error for the non-high vowels was small, between 0.3 and 0.6 dB, whereas the maximum error values for /u/ and /i/ were considerably greater at 1.0 and 1.8 dB, respectively. There would thus be only a small degree of distortion inherent in **EE** for the non-high vowels, with a somewhat greater distortion in **EE** likely for high vowels.

The **RA** estimates showed a maximum error of 1.3 percentage points for the non-high set of vowels (Figure 9). For /i/, however, the maximum error was more than 5 percentage points, whereas it was close to 3 percentage points for estimates from the /u/ vowel. Although numerically these errors may seem small, they amount to considerable distortion in the **RA** measure. Even a small variation in **RA** could have a noticeable impact on the acoustic properties of the source, as it is the relative change that is of main importance. For example, there was a 1.1 percentage points increase in the **RA** estimate between ***an*** = 0 (**RA** = 1.1%) and ***an*** = 10 mm^2 (**RA** = 2.2%) for /a/, which represents doubling (100% increase) of the **RA** value. This in turn would result in a 5.4 dB increase in the estimated attenuation in the source spectrum at 3 kHz (cf. Equation A2 in Appendix 1).

Figure 9 also shows a similar trend for estimates of the **RG** parameter. The maximum error was 4 percentage points for

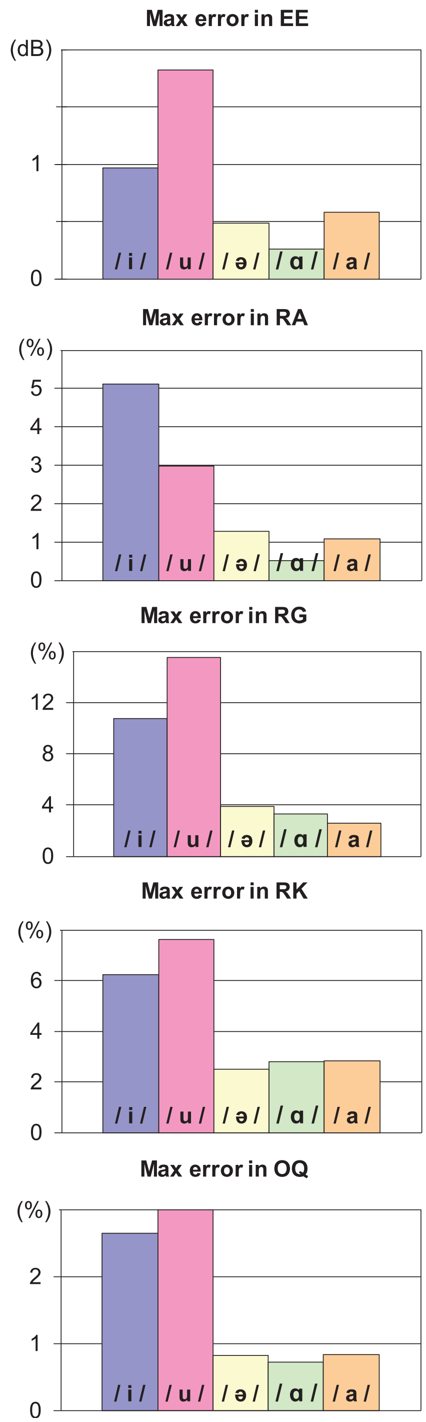

FIGURE 9. Summary of maximum errors observed in the estimates of five voice source parameters across various amounts of nasal coupling. Values for **EE** are in decibels, whereas values for **RA**, **RG**, **RK**, and **OQ** are in percentage points.

the non-high vowels, whereas for the high vowels, the maximum error was in the vicinity of 15 percentage points. Also for **RK** and **OQ**, the maximum errors were larger for high than for non-high vowels, although the errors were on the whole quite small. For **OQ**, the maximum errors for the non-high vowels were less than 1 percentage point, whereas for the high vowels, they were approximately 3 percentage points.

It is also important to consider the potential effects that the errors found in these parameters might have on the acoustic characteristics of the source spectrum. The **EE** value determines the overall amplitude of the whole spectrum (with the exception of the lowest components). **RG** together with **RK** determine the **OQ** value, which generally is strongly correlated with the relative amplitude levels of the first two harmonics of the source spectrum (H1 − H2). Here, we used the formula in the study by Fant[47(p. 128)] to convert OQ values for the LF model pulses to corresponding values of H1 − H2. Although H1 − H2 could of course be measured directly from the spectrum of the source signal, the computed values provide reasonable insight into the consequence of the changes in **OQ**. (For the range of values involved here, the maximum error should be less than 0.5 dB)[46]. **RA** mainly determines the slope of the source spectrum. Equation A2 was used to convert the **RA** value to the corresponding TL value to get an idea of how the amplitude level estimates are affected for relatively higher frequency source components.

In Figure 10, it can be seen that for the non-high vowels, the overall amplitude level is distorted by a maximum of 0.6 dB (**EE**), the distortion at the lower end of the spectrum is minimal at 0.2 dB (H1 − H2), whereas the estimated level at 3 kHz is as much as 6.3 dB lower than expected (TL). For the high vowels, the corresponding values are 1.8 dB (**EE**), 0.7 dB (H1 − H2), and 13 dB (TL).

It is clear from these values that the most crucial distortion of the source spectrum is caused by errors in the estimates of **RA**. When the vowel is nasalized, the **RA** estimates tend to increase: this exaggerates the amount of dynamic leakage and gives the impression that the downward slope of the source spectrum is greater than it really is. The general trend is that the error in **RA** increases with increasing nasal aperture, but this is not necessarily always the case, as can be seen clearly in the data in Figure 4, particularly for the vowels /u/, /a/, and /ɑ/.

DISCUSSION

The present study examines the effect of nasal coupling on the accuracy of glottal source measures obtained through inverse filtering based on an all-pole assumption of the vocal tract. HLsyn, a speech synthesizer based on a quasi-articulatory model of speech production, was used to generate a series of vowels with varying degrees of nasal coupling. The synthesized vowels were inverse filtered using standard inverse filtering techniques that did not specifically accommodate the acoustical effect of nasal coupling. The resultant source signals were subsequently modeled using the LF model, and a series of measures were made that estimate important acoustical determinants of the glottal source: excitation strength (**EE**), dynamic leakage (**RA**), normalized glottal frequency (**RG**), glottal pulse symmetry (**RK**), and open quotient (**OQ**). The estimates were then compared across nasal coupling conditions to establish the stability of these estimates when inverse filtering of nasalized utterances is accomplished in this way.

The results suggest that for non-high vowels, the use of inverse filtering to estimate LF source parameters in speech

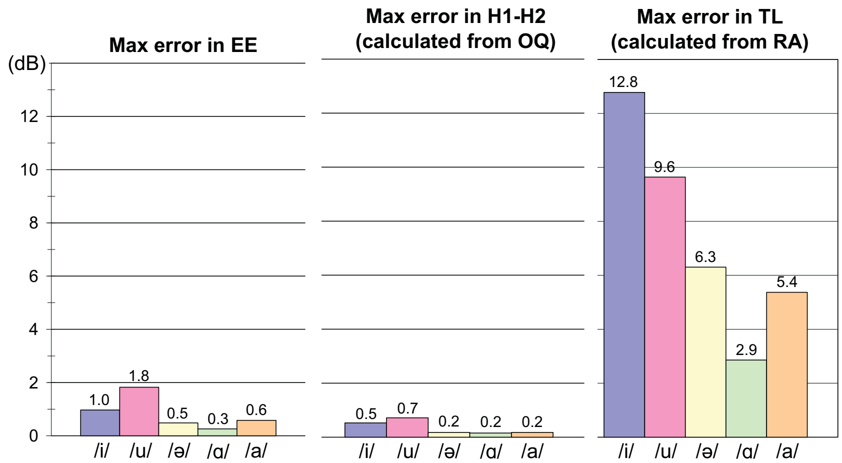

FIGURE 10. Maximum error for **EE** and calculated errors for H1 − H2 and TL.

that is moderately nasalized should yield reasonable estimates of parameters related to **EE**, **RG**, **RK**, and **OQ**. The errors for these parameters are likely to be negligible. The effects on the dynamic leakage, **RA**, is more striking: the results found here show **RA** estimates that are too high, and that consequently, the amount of dynamic leakage and the downward slope of the source spectrum are likely to be exaggerated.

For high vowels, the errors appear to be greater. Nevertheless, for **EE**, **RG**, **RK**, and **OQ**, the errors found are still relatively small and can probably be disregarded. Although the errors are larger than the corresponding errors for the non-high vowels, they are probably smaller than the typical errors one would get when analyzing (non-nasalized) natural speech. It is worth pointing out that the lower F_1 associated with the high vowels poses challenges to inverse filtering even without nasal coupling, particularly when F_1 comes into close proximity to f_0 (see further comments in Ref.[61]).

The errors in the **RA** parameter are, however, considerably larger for the high vowels: one would thus be led to believe that the source spectral slope in nasalized high vowels is substantially steeper than it actually is. From the results, it seems clear that the extent to which glottal source estimates are affected is related to the specific vocal tract configuration and particularly to tongue height.

The difference between vowels can be related to two factors. The first, and probably the main factor, is the extent to which the frequencies of the nasal pole and zero diverge. When they are close in frequency, they have minimal effect on the spectrum, and the more they separate, the greater the influence on the spectrum (making all-zero inverse filtering more problematic). In HLsyn, the increase in the frequency separation of the pole-zero pair with increasing nasal coupling depends on vowel quality: for any given size of the nasal aperture, the frequency separation is the greatest for /i/ followed by /u/. The second factor concerns the frequency distance between $F1$ (and sometimes $F2$) and the nasal pole (FN). When the nasal pole is relatively close to $F1$ (and/or $F2$), the effect can be more readily compensated for by adjusting some of the parameters of the all-zero filter. For /i/ in particular, the nasal pole occurs far from both $F1$ and $F2$.

The reason for the wider separation of the (first) pole-zero pair for nasalized high vowels is in part caused by the higher oral tract impedance at low frequencies for these vowels. As the sound transmission through the oral cavity is more restricted, the relative contribution of the nasal output to the overall speech output is greater. The frequency of the zero of the *oral* output is generally considerably lower than that of the zero of the *nasal* output.[20(p151)] The greater the contribution of the nasal output to the overall speech waveform, the closer the zero (in the spectrum of the speech output) will be to the zero of the nasal output. Consequently, the frequency of the zero will increase with increasing velopharyngeal opening, and for a given size of the opening, the frequency of the zero will tend to be relatively higher for vowels where the sound transmission through the oral cavity is more restricted (ie, for high vowels).

The degree of velopharyngeal opening in nasalized vowels depends on vowel height. Both X-ray data and electromyographic recordings of the elevator muscle of the soft palate (levator veli palatini) have revealed a relatively smaller opening in high vowels.[62,63] Similarly, Moll[64] found less velopharyngeal opening for high vowels than low vowels (see also comments in Ref.[20[pp156–159]]). This would suggest that the high vowels may be acoustically more sensitive to small variations in nasal coupling, with this increased sensitivity impacting both on the acoustic effect of nasal coupling and the resulting nasalized percept.[25] Therefore, as high vowels are generally produced with a smaller velopharyngeal opening, this (when analyzing natural speech data) could possibly counteract the larger errors in the source measures found in our synthesis-based measurements.

Contrary to expectation, some of the parameters showed rather abrupt changes as the ***an*** parameter increased. For example, it was noted that there was a drop in **EE** as the nasal aperture increases from 0 to 10 mm^2 for /u/, and to a lesser extent /i/ (Figure 4). The **EE** estimate is of course affected by the uncancelled pole-zero pair as ***an*** increases, but one would expect the effect to be very small for these low ***an*** values. One can only speculate as to the reason for this drop: it may perhaps be related to the fact that **EE** is only indirectly controlled in the

KLGLOTT88 model through the AV parameter; it may be that the **EE** value of the KLGLOTT88 model is slightly affected by other parameter settings of the synthesizer. The big picture is however clear: the variation in **EE** because of increasing nasal coupling is, on the whole, relatively small.

As can be seen in Figure 6, there is an abrupt change in the **RG** variation for /i/ going from $an = 25\ \mathrm{mm}^2$ to $30\ \mathrm{mm}^2$. There is also quite a marked drop in **RG** values for /u/ and /i/ at $an = 15\ \mathrm{mm}^2$. (Similar discontinuities can also be found for **RK** and **OQ** as the definitions of the three parameters are inextricably linked.) The explanation for this variation is most likely connected to the uncancelled oscillation of the nasal pole. It should be pointed out here that for all vowels except /a/, BNP is high (200 Hz) up to $an = 10\ \mathrm{mm}^2$. At $an = 15\ \mathrm{mm}^2$, the bandwidth begins to drop, and at $an = 20\ \mathrm{mm}^2$, it is 120 Hz. At $an = 25\ \mathrm{mm}^2$, it begins to increase again, rising linearly to 280 Hz at $an = 100\ \mathrm{mm}^2$. As the bandwidth is reduced, the distorting effect of the oscillation on the glottal pulse shape is increased. Because the amplitude of the oscillation is decreasing over the course of the glottal cycle (at a rate determined by its bandwidth), its impact is at its greatest at the beginning of the pulse, at the time point of glottal opening, t_o. Consequently, the tuning of t_o may have to be changed to optimize the source modeling. The nature of the change, however, is difficult to predict, because it depends on the frequency of the nasal pole (which is increasing with increasing nasal aperture) and which will ultimately determine the precise phase relationship between the oscillation of the nasal pole and the time point of glottal opening. Depending on the phase relationship that results, t_o may have to be adjusted to be either slightly earlier or slightly later. The reason the effect is much less noticeable in the non-high vowels is because of the much smaller frequency difference between the nasal pole-zero pair for these vowels. Note again that as an is increasing (above $an = 25\ \mathrm{mm}^2$), BNP is increasing, and its influence is reduced.

It should be clear from the above discussion that the interactions between the many parameters involved are highly complex, and one could expect seemingly arbitrary changes to occur. The main point, however, is that the overall effect of the variation in nasal aperture on **RG**, **RK**, and **OQ** is relatively small.

It needs to be pointed out that these findings are of course dependent on the characteristics of the synthesizer. There are a number of simplifications in the modeling of the nasal tract in HLsyn, which could potentially affect the optimization of the inverse filtering and hence the accuracy of the parameterization. The acoustic effect of the nasal tract is modeled using only a single pole-zero pair. Higher modes, the acoustic coupling to the sinus cavities, and the possibility of asymmetries in the left and right nasal channels are not taken into account. Nevertheless, we would not expect these simplifications to affect our results in any major way. In fact, our results suggest that it is the oral tract characteristics that are critical to the accuracy of the analysis, rather than the precise details of the nasal tract transfer function. A further simplification concerns the one-dimensional modeling of the wave propagation in the vocal tract: cross-modes are not considered. The effects of cross-modes are likely to be speaker specific and highly variable. Therefore, it is difficult to predict how they would affect the inverse filtering optimization.

Furthermore, one should bear in mind that the results presented here were obtained under essentially "ideal conditions," that is, they represent the errors introduced when the inverse filtering has been done optimally. The analysis carried out involved detailed interactive inverse filtering, where all the settings of the inverse filter were manually optimized (within the constraints outlined earlier) so as to achieve the best possible cancellation of the effects of the vocal tract filter. Often automatic inverse filtering techniques are used, typically based on linear predictive coding. These techniques are known to be error prone when analyzing natural speech data, even for non-nasalized speech.[44] When using such techniques, the inverse filtering is likely to be less than optimal, and therefore the additional errors introduced in the source estimates for nasalized vowels are likely to be less predictable.

CONCLUSIONS

To summarize, the current findings suggest that when analyzing nasalized vowels using inverse filtering based on an all-pole assumption of the vocal tract, the effect on the **EE**, **RG**, **RK**, and **OQ** parameters is relatively small, regardless of the degree of nasal coupling. These findings further suggest that it is reasonable to estimate these parameters in populations for which there is even a large degree of nasal coupling, such as those with dysarthria or hearing loss. The **RA** parameter—a measure relating to the amount of dynamic leakage in the return phase of the glottal cycle and a main determinant of the slope of the voice source spectrum—can be adversely affected by nasal coupling, particularly for high vowels.

The findings presented here are of course predicated on the characteristics of the synthesizer used to simulate varying degrees of nasal coupling. Although the HLsyn model simulates the main characteristics of nasality in a way that is consistent with our current understanding, these findings need to be verified using real speakers. Future research might involve human speakers simulating varying degrees of nasality using a specially constructed prosthetic obturator that would carefully control the nasal aperture dimensions.[65] This approach would also require monitoring of vocal fold behavior, perhaps through electroglottography or high-speed imaging, to ensure some degree of consistency across varying degrees of nasal coupling. Although we would not have direct access to the voice source signal, knowing the degree of nasal coupling would facilitate inverse filtering based on a more realistic vocal tract model, while parameters derived from the vocal fold monitoring would provide independent evidence of consistent vocal fold behavior. The results from such inverse filtering could then be directly compared with the results from inverse filtering based on the simple all-pole assumption of the vocal tract.

Acknowledgments

The authors thank Corine Bickley and Helen Hanson for their comments on earlier drafts of this manuscript.

REFERENCES

1. Alku P, Vilkman E. Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by inverse filtering. *Speech Commun*. 1996;18:131–138.
2. Gobl C. Voice source dynamics in connected speech. In: *STL-QPSR*, 1988; 29. Stockholm, Sweden: Speech, Music and Hearing KTH; 1988:123–159. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1988/1988_29_1_123-159. Accessed May 29, 2012.
3. Gobl C. A preliminary study of acoustic voice quality correlates. In: *STL-QPSR*, 1989;30. Stockholm, Sweden: Speech, Music and Hearing KTH; 1989:9–21. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1989/1989_30_4_009-022.pdf. Accessed May 29, 2012.
4. Gobl C, Ní Chasaide A. Acoustic characteristics of voice quality. *Speech Commun*. 1992;11:481–490.
5. Gobl C, Ní Chasaide A. Voice source variation in the vowel as a function of consonantal context. In: Hardcastle WJ, Hewlett N, eds. *Coarticulation: Theory, Data and Techniques*. Cambridge, MA: Cambridge University Press; 1999:122–143.
6. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J Acoust Soc Am*. 1988;84:511–529.
7. Karlsson I. Modelling voice variations in female speech synthesis. *Speech Commun*. 1992;11:491–495.
8. Ní Chasaide A, Gobl C. Contextual variation of the vowel voice source as a function of adjacent consonants. *Lang Speech*. 1993;36(pt 2–3):303–330.
9. Pierrehumbert J. A preliminary study of the consequences of intonation for the voice source. In: *STL-QPSR*, 1989;30. Stockholm, Sweden: Speech, Music and Hearing KTH; 1989:23–36. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1989/1989_30_4_023-036.pdf. Accessed May 29, 2012.
10. Rothenberg M. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *J Acoust Soc Am*. 1973;53:1632–1645.
11. Yanushevskaya I, Gobl C, Kane J, Ní Chasaide A. An exploration of voice source correlates of focus. Presented at: *Interspeech 2010*; 2010; Makuhari, Japan:462–465.
12. Yanushevskaya I, Tooher M, Gobl C, Ní Chasaide A. Time- and amplitude-based voice source correlates of emotional portrayals. In: Paiva ACR, Prada R, Picard RW, eds. *Affective Computing and Intelligent Interaction*, 2007;Vol 4738. Berlin, Heidelberg: Springer; 2007:159–170. Available at: http://www.springerlink.com/index/10.1007/978-3-540-74889-2_15. Accessed May 29, 2012.
13. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: an experimental framework and initial results. *J Speech Hear Res*. 1989;32:373–392.
14. Holmberg EB, Doyle P, Perkell JS, Hammarberg B, Hillman RE. Aerodynamic and acoustic voice measurements of patients with vocal nodules: variation in baseline and changes across voice therapy. *J Voice*. 2003;17: 269–282.
15. Kane P, Ní Chasaide A. A comparison of the dysphonic and normal voice source. *J Clin Speech Lang Stud*. 1992;1:17–29.
16. Mahshie J, Öster AM. Velopharyngeal function: an electromyographic study. In: *STL-QPSR*, 1991;11. Stockholm, Sweden: Speech, Music and Hearing KTH; 1991:19–27. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1991/1991_32_2-3_019-027.pdf. Accessed May 29, 2012.
17. Sundberg J, Gauffin J. Waveform and spectrum of the glottal voice source. In: *STL-QPSR*, 1978;19. Stockholm, Sweden: Speech, Music and Hearing KTH; 1978:35–50. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1978/1978_19_2-3_035-050.pdf. Accessed August 29, 2012.
18. Gauffin J, Sundberg J. Spectral correlates of glottal voice source waveform characteristics. *J Speech Hear Res*. 1989;32:556–565.
19. Arroabarren I, Carlosena A. Inverse filtering in singing voice: a critical analysis. *IEEE Trans Audio Speech Lang Process*. 2006;14:1422–1431.
20. Fant G. *Acoustic Theory of Speech Production*. The Hague, The Netherlands: Mouton De Gruyter; 1970. 1st ed. 1960.
21. Stevens KN. *Acoustic Phonetics*. Cambridge, MA: MIT Press; 1998.
22. Lindqvist-Gauffin J, Sundberg J. Acoustic properties of the nasal tract. *Phonetica*. 1976;33:161–168.
23. Maeda S. The role of the sinus cavities in the production of nasal vowels. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing.* Vol 1. Paris, France: IEEE; 1982:911–914. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1171561. Accessed August 29, 2012.
24. Delattre P. The acoustical properties of nasal vowels and consonants. *Studia Linguistica*. 1954;8:103–109.
25. House AS, Stevens KN. Analog studies of the nasalization of vowels. *J Speech Hear Disord*. 1956;21:218–232.
26. Hawkins S, Stevens KN. Acoustic and perceptual correlates of the non-nasal–nasal distinction for vowels. *J Acoust Soc Am*. 1985;77:1560–1575.
27. Stevens KN, Fant G, Hawkins S. Some acoustical and perceptual correlates of nasal vowels. In: Channon R, Shockey L, eds. *Festschrift in honor of Ilse Lehiste*. Dordrecht, Holland: Foris; 1987:241–254.
28. Feng G, Castelli E. Some acoustic features of nasal and nasalized vowels: a target for vowel nasalization. *J Acoust Soc Am*. 1996;99:3694–3706.
29. Karlsson I, Liljencrants J. Diverse voice qualities: models and data. In: *TMH-QPSR*, 1996;37. Stockholm, Sweden: Speech, Music and Hearing KTH; 1996:143–146. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1996/1996_37_2_143-146.pdf. Accessed May 29, 2012.
30. Ananthapadmanabha TV. Acoustic analysis of voice source dynamics. In: *STL-QPSR*, 1984;25. Stockholm, Sweden: Speech, Music and Hearing KTH; 1984:1–24. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1984/1984_25_2-3_001-024.pdf. Accessed May 29, 2012.
31. Thompson EC, Murdoch BE. Disorders of nasality in subjects with upper motor neuron type dysarthria following cerebrovascular accident. *J Commun Disord*. 1995;28:261–276.
32. Ackermann H, Ziegler W, Petersen D. Dysarthria in bilateral thalamic infarction. A case study. *J Neurol*. 1993;240:357–362.
33. Theodoros DG, Murdoch BE. Laryngeal dysfunction in dysarthric speakers following severe closed-head injury. *Brain Inj*. 1994;8:667–684.
34. Theodoros DG, Murdoch BE, Chenery HJ. Perceptual speech characteristics of dysarthric speakers following severe closed head injury. *Brain Inj*. 1994;8:101–124.
35. Hartelius L, Runmarker B, Andersen O. Prevalence and characteristics of dysarthria in a multiple-sclerosis incidence cohort: relation to neurological data. *Folia Phoniatr Logop*. 2000;52:160–177.
36. Nguyen LHP, Allegro J, Low A, Papsin B, Campisi P. Effect of cochlear implantation on nasality in children. *Ear Nose Throat J*. 2008;87:138. 140–143.
37. Leder SB, Spitzer JB. A perceptual evaluation of the speech of adventitiously deaf adult males. *Ear Hear.* 1990;11:169–175.
38. Austin SF. Movement of the velum during speech and singing in classically trained singers. *J Voice*. 1997;11:212–221.
39. Jennings JJ, Kuehn DP. The effects of frequency range, vowel, dynamic loudness level, and gender on nasalance in amateur and classically trained singers. *J Voice*. 2008;22:75–89.
40. Birch P, Gümoes B, Stavad H, Prytz S, Björkner E, Sundberg J. Velum behavior in professional classic operatic singing. *J Voice*. 2002;16:61–71.
41. Sundberg J, Birch P, Gümoes B, Stavad H, Prytz S, Karle A. Experimental findings on the nasal tract resonator in singing. *J Voice*. 2007;21:127–137.
42. Perkell JS, Holmberg EB, Hillman RE. A system for signal processing and data extraction from aerodynamic, acoustic, and electroglottographic signals in the study of voice production. *J Acoust Soc Am*. 1991;89(4 pt 1): 1777–1781.
43. Huffman MK. Measures of phonation type in Hmong. *J Acoust Soc Am*. 1987;81:495.
44. Gobl C, Ní Chasaide A. Voice source variation and its communicative functions. In: Hardcastle WJ, Laver J, Gibbon FE, eds. *The Handbook*

of Phonetic Sciences. Oxford, UK: Blackwell Publishing Ltd.; 2010: 378–423.
45. Fant G, Liljencrants J, Lin Q. A four-parameter model of glottal flow. In: *STL-QPSR*, 1985;26. Stockholm, Sweden: Speech, Music and Hearing KTH; 1985:1–13. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1985/1985_26_4_001-013.pdf. Accessed May 29, 2012.
46. Fant G. The LF-model revisited. Transformations and frequency domain analysis. In: *STL-QPSR*, 1995;36. Stockholm, Sweden: Speech, Music and Hearing KTH; 1995:119–156. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1995/1995_36_2-3_119-156.pdf. Accessed May 29, 2012.
47. Fant G. The voice source in connected speech. *Speech Commun*. 1997;22: 125–139.
48. Gobl C. *The voice source in speech communication—production and perception experiments involving inverse filtering and synthesis* [PhD thesis]. Stockholm, Sweden: Royal Institute of Technology (KTH); 2003.
49. Stevens KN, Bickley CA. Constraints among parameters simplify control of Klatt formant synthesizer. *J Phon*. 1991;19:161–174.
50. Hanson HM, Stevens KN. A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *J Acoust Soc Am*. 2002;112(3 pt 1):1158–1182.
51. Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am*. 1990;87: 820–857.
52. McGowan RS, Hanson HM, Stevens KN, Gau DW. M4ELD1 Talker transformation through synthesis. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*. Vol 1. San Francisco, California: Berkeley: University of California; 1999:137–140.
53. Ogden R, Hawkins S, House J, et al. ProSynth: an integrated prosodic approach to device-independent, natural-sounding speech synthesis. *Comput Speech Lang*. 2000;14:177–210.
54. Mahshie J, Bickley C. Production of voiced ingressive plosives by deaf speakers: an analysis by synthesis. In: *Proceedings of the XVth International Congress of Phonetic Sciences*. Barcelona, Spain: Barcelona: Universitat Autonoma de Barcelona; 2003:3257–3260.
55. Bickley C, Bruckert E. Improvements in the voice quality of DECtalk®. In: *Proceedings of the 2002 IEEE Workshop on Speech Synthesis*. Santa Monica, CA: IEEE; 2002:55–58. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1224372&tag=1. Accessed October 12, 2012.
56. Warren DW, Devereux JL. An analog study of cleft palate speech. *Cleft Palate J*. 1966;3:103–114.
57. Moon JB, Weinberg B. Two simplified methods for estimating velopharyngeal orifice area. *Cleft Palate J*. 1985;22:1–10.
58. Gobl C, Ní Chasaide A. Techniques for analysing the voice source. In: Hardcastle WJ, Hewlett N, eds. *Coarticulation: Theory, Data and Techniques*. Cambridge, MA: Cambridge University Press; 1999:300–320.
59. Gobl C, Monahan P, Ní Chasaide A. Intrinsic voice source characteristics of selected consonants. In: *Proceedings of the XIIIth International Congress of Phonetic Sciences*. Vol 1. Stockholm, Sweden: Stockholm, Royal Institute of Technology (KTH) and Stockholm University; 1995: 74–77.
60. Fant G. Glottal source and excitation analysis. In: *STL-QPSR*, 1979;20. Stockholm, Sweden: Speech, Music and Hearing KTH; 1979:85–107. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1979/1979_20_1_085-107.pdf. Accessed May 29, 2012.
61. Alku P, Story B, Airas M. Estimation of the voice source from speech pressure signals: evaluation of an inverse filtering technique using physical modelling of voice production. *Folia Phoniatr Logop*. 2006;58: 102–113.
62. Lubker JF. An electromyographic-cinefluorographic investigation of velar function during normal speech production. *Cleft Palate J*. 1968;5: 1–18.
63. Lubker J, Liljencrants J, Fritzell B, Lindqvist-Gauffin J. Electroglottographic and glottal air flow measurements for deaf and normal-hearing speakers. In: *STL-QPSR*, 1970;32. Stockholm, Sweden: Speech, Music and Hearing KTH; 1970:9–20. Available at: http://www.speech.kth.se/prod/publications/files/qpsr/1970/1970_11_4_009-020.pdf. Accessed May 29, 2012.
64. Moll KL. Velopharyngeal closure on vowels. *J Speech Hear Res*. 1962;5: 30–37.
65. Kent L. The effects of oral-to-nasal coupling on the perceptual, physiological, and acoustical characteristics of vowels [unpublished doctoral dissertation]: Iowa City, IA: University of Iowa; 1966.

APPENDIX 1

Deriving normalized RA values

Due to its exponential nature, the effect of the LF model return phase on the source spectrum is approximately that of a first-order low-pass filter. The cutoff frequency, *FA*, of the filter is

$$FA = \frac{1}{2\pi\ \mathrm{T_a}} = \frac{f_0}{2\pi\ \mathbf{RA}}\quad (\mathrm{Hz}) \tag{A1}$$

where T_a is an approximate measure of the time constant of the exponential function (Figure 1). The TL value determines the additional attenuation, in decibels, of the source spectrum at 3000 Hz. Therefore, TL can be expressed as a function of *FA* (and **RA**) as follows:

$$\begin{aligned}\mathrm{TL}(FA) &= 10\cdot\log_{10}\left(1+\left(\frac{3000}{FA}\right)^2\right)\\ &= 10\cdot\log_{10}\left(1+\left(\frac{6000\pi\ \mathbf{RA}}{f_0}\right)^2\right)\quad(\mathrm{dB})\end{aligned} \tag{A2}$$

From Equation A2, it is clear that the corresponding (linear) amplitude change is determined by the scale factor tl_{lin} as indicated in Equation 3:

$$\begin{aligned}tl_{lin}(FA) &= \left(1+\left(\frac{3000}{FA}\right)^2\right)^{-0.5}\\ &= \left(1+\left(\frac{6000\pi\ \mathbf{RA}}{f_0}\right)^2\right)^{-0.5}\end{aligned} \tag{A3}$$

Increasing TL by ΔTL (in decibels), produces a further amplitude reduction of $10^{-\Delta\mathrm{TL}/20}$ in the spectrum at 3000 Hz. The linear scale factor determining the total amplitude changed caused by TL + ΔTL is shown on the left-hand side of Equation A4, which equals the expression on the right-hand side, where **RA** in Equation A3 has been replaced by the normalized **RA** value, here denoted $\mathbf{RA}_{norm}$.

$$\begin{aligned}&10^{-\frac{\Delta\mathrm{TL}}{20}}\left(1+\left(\frac{6000\pi\ \mathbf{RA}}{f_0}\right)^2\right)^{-0.5}\\ &= \left(1+\left(\frac{6000\pi\ \mathbf{RA}_{norm}}{f_0}\right)^2\right)^{-0.5}\end{aligned} \tag{A4}$$

By solving Equation A4 for $\mathbf{RA}_{norm}$, we arrive at the expression shown in Equation A5. Equation A5 allows us to calculate the normalized **RA** value given a positive TL change, ΔTL (in decibels), from the original **RA** value and f_0.

$$\mathbf{RA}_{norm}(\Delta\mathrm{TL}) = \mathbf{RA}\cdot 10^{\frac{\Delta\mathrm{TL}}{20}} \times\sqrt{1+\left(\frac{f_0}{6000\pi\ \mathbf{RA}}\right)^2\left(1-10^{\frac{-\Delta\mathrm{TL}}{10}}\right)} \quad \text{(A5)}$$

For example, to normalize a TL value of 8 dB to correspond to TL = 5, a ΔTL value of -3 is used in Equation A5, which will result in a scaling of the original **RA** values by a factor of 0.68.

APPENDIX 2

The relationship between 'LF parameters' and 'standard *JoV* parameters'

Voice source studies published in *Journal of Voice (JoV)* typically use a specific set of source parameters. This *JoV* set of parameters includes MFDR (maximum flow declination rate), Q_{open} (open quotient), Q_{closed} (closed quotient), AC_{amp} (pulse amplitude), NAQ (normalized amplitude quotient = $(AC_{amp}/\mathrm{MFDR})/T_0$), H1 – H2 (difference in amplitude level between the first two harmonics of the source spectrum), and SQ (speed quotient, the ratio of the opening and closing times of the glottal pulse).

The definitions and acronyms for most of these parameters are different from those we use in this article. Therefore, we briefly outline the relationship between the two sets of parameters. **EE** and MDFR are effectively equivalent. However, the amplitude of **EE** is determined at the *maximum discontinuity* in the glottal pulse: this point in the waveform typically coincides with the point of maximum flow declination rate, with the exception, perhaps, for very breathy voicing, where the source signal becomes close to sinusoidal. **OQ** (open quotient) is equivalent to Q_{open} and $1 - Q_{closed}$. **RK** (glottal skew) is equivalent to 1/SQ. In case of **RG** (the normalized glottal frequency) there is no direct equivalent in the standard *JoV* set. However, variation in **RG** is negatively correlated with H1 – H2, as a higher **RG** will shift the emphasis from the first harmonic towards the second harmonic in the source spectrum, thus reducing H1 – H2. For **RA** (the normalized effective duration of the return phase) there is no a matching parameter in the *JoV* set, as the principal effect of a change in **RA** is on the amplitude levels of the higher harmonics. However, in terms of the LF model, an increase in **RA** will also result in an increase in AC_{amp} and NAQ, and if **RG** is low enough to boost the first harmonic, H1 – H2 will increase as well.