

Laryngeal and Aerodynamic Adjustments for Voicing Versus Devoicing of /h/: A Within-Speaker Study

*[†]Laura L. Koenig, †Jorge C. Lucero, and *W. Einar Mencl, *New Haven, CT, †Brooklyn, NY, and ‡Brasilia, Brazil

Summary: The aim of this within-speaker case study was to explore how effectively a phonetically trained speaker could alter the likelihood of voicing around abduction, and what changes he made to do so. An American English-speaking male produced intervocalic /h/ in varying loudness and vowel contexts. When given no specific instructions about voicing (block 1), he produced almost entirely voiced /h/. He was then asked to devoice /h/ (block 2). Measures of voicing, baseline airflow, pulse amplitudes, fundamental frequency (F0), open quotient, and speed quotient were made from oral airflow signals. Subglottal pressure was estimated from intraoral pressures during /p/. In block 2, the speaker produced 70% devoiced /h/. He achieved this by making several changes associated with higher phonation threshold pressures: greater abduction degrees, lower subglottal pressures, greater longitudinal tension of the vocal folds, and altered laryngeal settings. Qualitative inspection of the DC flow contours along with correlational and principal components analyses indicated widespread changes in respiratory, laryngeal, and supralaryngeal settings, and differing interrelationships among variables. Our speaker showed tacit knowledge of the range of parameters affecting voicing. Differing relationships among variables across the two blocks support a view of phonation as a dynamic process, where speakers adjust multiple parameters, simultaneously.

Key Words: Phonation–Laryngeal aerodynamics–Within-speaker study.

INTRODUCTION

This case study forms part of a larger cross- and within-speaker investigation into voicing behavior around abduction, focusing specifically on the consonant /h/ in English. Although English /h/ is phonologically classified as voiceless, several authors have noted that voiced allophones may occur, especially in intervocalic contexts.^{1–3} Our strategy is to consider how the presence or absence of phonation in /h/, or the length of a devoiced interval, varies as a function of parameters known to affect voicing. As a context for exploring voicing control, /h/ is particularly useful for two reasons: (1) because /h/ does not involve an upper vocal tract constriction, laryngeal and aerodynamic conditions can be monitored noninvasively throughout abduction by means of oral airflow signals. (2) Because we are doing modeling as well as measurement, it is practical first to consider phonation in the simple case of abduction without supralaryngeal constriction. Once the model has been refined to fit these data, obstruent consonants can be added.

Our general questions are these: (1) Can we account for why some speakers produce mostly fully-voiced /h/, whereas others often produce voiceless /h/? (2) Can we account for the difference between voiced versus devoiced productions of /h/ within a single subject? In general, we predict that individuals who produce more devoiced /h/ will show laryngeal and aerodynamic characteristics associated with higher phonation threshold pressures (see [discussion below](#)), but whether some adjustments are more common than others is unknown. Similarly, within speakers, we expect devoicing to correlate with

factor settings that increase phonation threshold pressures, but speakers may differ in which factors they manipulate.

An issue relevant to both the cross- and within-speaker analyses is the degree to which a person can manipulate his/her voicing behavior. Although we suspect that an individual's typical phonatory behavior reflects, to some degree, inherent aspects of his/her production system, speakers can obviously vary fundamental frequency (f0) and voice quality over a certain range, and thus presumably also have the physiological ability to adjust the likelihood of phonation. The purpose of this study was to explore how effectively a single speaker could alter his /h/ voicing behavior, and, if he could, what changes he made to do so.

Literature

Sustained vocal fold vibration (voicing) occurs when a speaker achieves an appropriate balance among transglottal pressure, vocal fold thickness and longitudinal tension, degree of abduction, glottal configuration, and tissue damping.^{4–9} The phonation threshold pressure,⁸ which represents the minimum transglottal pressure required for phonation, provides a convenient way of quantifying the conditions under which voicing is possible. All else being equal, phonation threshold pressures are lower (ie, phonation can be initiated or sustained with less of a driving force) for greater vocal fold thicknesses, lower longitudinal tensions, smaller abduction degrees, more convergent glottal angles, and lower tissue damping values.

Our work on /h/ voicing in American English indicates that some speakers regularly phonate throughout abduction for /h/ in an intervocalic, stressed, word-initial position,¹⁰ and further that /h/ voicing in this context may be more common in men than in women or children. This is consistent with results showing that men tend to have longer, thicker, less stiff, more dense vocal folds, along with more complete glottal closure as a result of greater tissue bulk in the vocal folds.⁷ All of these factors should lead men, on average, to have a greater likelihood of voicing given the same transglottal pressure.^{4,8,11} Our modeling

Accepted for publication April 5, 2007.

Preliminary results of this study were presented at the 150th Meeting of the Acoustical Society of America, Minneapolis, MN.

From the *Haskins Laboratories, New Haven, Connecticut; †Long Island University, Department of Communication Sciences and Disorders, Brooklyn, New York; and the ‡Department of Mathematics, University of Brasilia, Brasilia, Brazil.

Address correspondence and reprint requests to Laura L. Koenig, Haskins Laboratories, 300 George Street, New Haven, CT 06511. E-mail: koenig@haskins.yale.edu

Journal of Voice, Vol. 22, No. 6, pp. 709–720

0892-1997/\$34.00

© 2008 The Voice Foundation

doi:10.1016/j.jvoice.2007.04.001

work^{12,13} similarly indicates that scaling a simple laryngeal model down to sizes appropriate for women and children yields a smaller range of conditions that permit voicing.

The wide variety of parameters that can be adjusted to achieve voicing/devoicing suggests that individuals may arrive at phonatory states using speaker-specific combinations of laryngeal and aerodynamic adjustments. In support of this hypothesis, our study of voiced and voiceless /h/ in women¹⁴ showed that all six speakers had unique patterns of laryngeal, supralaryngeal, and aerodynamic conditions leading to devoicing. At the same time, the range of factors affecting phonation should allow individual speakers some latitude in the configurations they use in a given situation.

A general question that arises from our past work is the extent to which subject-specific patterns reflect intrinsic anatomical and physiological parameters versus settings over which speakers have some control. To address this question, in this study we explicitly asked a phonetically trained speaker to vary his /h/ voicing behavior. On discovering that he was quite successful at this task, we then investigated how he produced this change. This work supplements our cross-speaker analyses by providing insight into the nature of speaker differences we have observed in the past. In particular, this within-speaker analysis allows us to explore the degree to which a speaker may be able to use his/her range of phonatory possibilities. We attempt to characterize fully the laryngeal system of a speaker who, by virtue of his phonetic training, should have good awareness of how to manipulate voicing-related parameters for particular purposes. We assume that our speaker has a laryngeal system in which some parameters are essentially fixed, whereas others are more malleable. We see this as analogous to studies that explore the behavior of a particular laryngeal model, with some constant characteristics and some variable settings.

METHODS

The instrumentation and most measures used here are identical to our previous study.¹⁴ The new features introduced in this study are (1) the overt voicing manipulation; (2) measures of f_0 during preceding and following vowels; and (3) voice-source measures.

Speaker

Our main participant selection criterion was phonetic training. Because voicing of /h/ is not distinctive in English, naive (ie, untrained) speakers cannot be expected to vary /h/ voicing on request. At the same time, it was unclear to us how successful even a phonetician would be on such a task. We recruited a 42-year-old native American English-speaking male with linguistic training and extensive professional experience in speech and voice for the theater. He provided informed consent to participate in the study and was recorded in a single session.

Although our speaker was in good vocal health when we recorded him, his responses to our background questionnaire indicated a positive history of voice intervention, so we questioned him in some detail about his vocal/medical

background. He indicated that, approximately 6 years before the time of recording, he had a granuloma removed from the posterior, nonvibratory, cartilaginous portion of the left vocal fold. The granuloma had resulted from abrasion of the epithelial layer during a coughing fit caused by a nonchronic bronchial infection. Postsurgical longitudinal stroboscopic evaluation performed by the speaker's laryngologist indicated normal vocal fold appearance. At the time of recording, the speaker was using daily administration of mometasone furoate monohydrate (nasonex) supplemented with saline nasal washes to address chronic allergic rhinitis. The speaker indicated that his laryngologist believed the condition to be firmly under control, and his voice quality was within normal limits as judged by the first author. Also at the time of recording, he was being treated successfully for Hashimoto's Disease. Neither the thyroid dysfunction nor thyroid medication related to this condition would be expected to affect his voice in any way. In short, we did not feel that our speaker's medical history rendered him unsuitable for our present purposes.

Speech materials

The speaker was asked to produce multiple repetitions of three utterances: "A Poppa Hopper," "A Poppa Hippie," and "A Poppa Hooper." Thus, the target /h/ was always intervocalic, initiating a stressed syllable, and flanked by the bilabial stop /p/ to allow recording of intraoral pressure peaks in the vicinity of /h/. The vowel variation (i a u) in the /h/-initiated syllable was intended to induce slight variation in supraglottal resistance, intrinsic f_0 , and possibly voice quality.^{15,16} Utterances were presented in randomized order five times in each of three blocked loudness conditions of normal (N), loud (L), and soft (S) speech. This manipulation was included to ensure a range of subglottal pressures over the recording session. The speaker was specifically asked not to whisper in the soft condition. Utterances were presented verbally in normal, loud, and soft volumes to encourage compliance with the loudness variation, and the speaker repeated each utterance five times per presentation.

Manipulation of /h/ voicing

The speaker was first introduced to the stimulus materials and loudness conditions. Then, for the first half of the session (block 1), he was recorded without further instructions; that is, he was not asked to attend to his /h/ voicing characteristics. Online monitoring indicated that he phonated through almost all tokens of /h/. At the end of this self-selected condition, the experimenter pointed out that the speaker's /h/ productions had mostly been voiced, and asked him to try, in another recording block, to produce devoiced /h/. He was permitted a few minutes of practice, with verbal feedback, and then recording block 2 began, identical to the first except for a different utterance randomization order. Once the second block began, no additional feedback on /h/ voicing was provided. Combining block 1 (speaker-selected) and block 2 (targeted devoicing), 452 tokens of /h/ were available for analysis. These represent two voicing blocks \times three utterances \times three loudness conditions \times five presentations \times five repetitions per presentation (plus two extra repetitions when our speaker miscounted).

Signals

Three signals were recorded simultaneously for all utterances: (1) sound pressure (ie, a microphone signal); (2) oral airflow (using a pneumotachograph (Glottal Enterprises, Syracuse, NY) and hardware MSIF); and (3) intraoral pressure, collected using a catheter transducer (CT/S) (Gaeltek, Medical Measurement Inc., Hackensack, NJ) affixed within a piece of medical tubing fed through the airflow mask to rest between the speaker's lips during bilabial closures, with the length adjusted so that the tube did not interfere with lingual movement. Sound pressure was sampled at 20 kHz; airflow and intraoral air pressure were sampled at 10 kHz. After the recording session, airflow signals were calibrated using a rotameter, and pressure signals were calibrated using a water manometer. Because the pressure transducer showed some drift over the course of the recording session, a unique offset was established for each input set (five repetitions of an utterance) by taking an average over a short region during a flat pressure region before or after the utterances, if possible, or else during one of the stressed vowels, and setting this to zero. The pressure peak values thus represent the increase during bilabial closure relative to this baseline.

Signal processing

The sound pressure signal was used to verify that the utterances were perceptually adequate renditions of the intended stimuli. All measures were made from the flow or pressure signals or signals derived from them. First, the flow was lightly smoothed with a five-point triangular window to eliminate high-frequency noise. Then, both the flow and pressure signals were smoothed with a wide triangular window (133 points) until all or most evidence of vocal fold vibration was obliterated. An AC flow signal was obtained by subtracting the slowly varying DC (smoothed) flow from the original flow signal. Finally, first derivative (velocity) signals were obtained from the smoothed flow and pressure, using a three-point difference algorithm, and the resultant was again smoothed with a 133-point window so that zero-crossings could be obtained easily. Zero-crossings in these smoothed velocity signals were used to label the peak pressures in /p/ and peak flows in /h/, as described below.

Measures

Our measures were designed to sample across the full range of factors that affect voicing thresholds. Specifically, we sought information on subglottal pressure, vocal fold abduction, degree of longitudinal tension, and general laryngeal setting. Figure 1 shows an example of the airflow and smoothed intraoral pressure signals for a whole utterance (panel A), examples of devoiced (panel B) and voiced (panel C) productions of /h/, and an expansion of the AC flow signal in the region around the voicing break in a devoiced token (panel D). Panels B and C include both the original (thin lines) and smoothed (thick lines) flow signals. The following measures were made for each token:

- (1) Peak intraoral pressures in the two /p/s surrounding the /h/ were obtained using the zero-crossings in the pressure velocity signal that corresponded to the pressure maxima.

As shown in the example in Figure 1, the pressure in /p/ typically rose quickly to a maximum value. These two maximum values were averaged and used as an estimate of subglottal pressure during the utterance.^{17,18}

- (2) Analogously, the peak baseline (DC or smoothed) flow was determined for each /h/, using zero-crossings in the smoothed flow velocity signal. In an open vocal tract, the DC airflow at the mouth is approximately proportional to the airflow at the glottis^{6,19} (neglecting any rotational flows that may arise from airstream jet formation in the vocal tract). Airflow at the glottis, in turn, varies as a function of glottal area (ie, abduction). Thus, the maximum DC flow value provides an indirect measure of the degree of glottal opening for a given token. Below, this maximum flow value is referred to as hFlowMax. (In our previous study,¹⁴ this measure was referred to as hPk.)
- (3) Times of voicing offset and onset in /h/ were obtained visually. We found previously¹⁴ that these visually defined measures showed high intrarater reliability ($r = 0.95$, $P < 0.001$, and mean differences of less than 1 millisecond). Two durations were then calculated: time between voicing offset and hFlowMax (VOffTh) and time between hFlowMax and voicing onset (VOnTh). In devoiced tokens, VOnTh is analogous to Voice Onset Time (VOT) in an aspirated stop, where the timing of peak glottal opening is correlated with the time of oral release.²⁰ In cases of fully voiced /h/, both VOnTh and VOffTh were approximately zero (cf. Figure 1, panel C), that is, less than the duration of a single period. (Because our pulse-by-pulse measurement routines required a whole number of glottal pulses before and after the maximum flow peak in /h/, pulses whose open phase overlapped with the hFlowMax label were not measured.) To allow a binary distinction between voiced and devoiced tokens, a two-pulse-period criterion was defined, based on the speaker's average f_0 measures before and after /h/. Productions with a voicing break greater than the duration of two pitch periods were considered to be devoiced.
- (4) The DC flow amplitudes were measured at the times of voicing offset and onset (DCOff, DCon). In devoiced cases, these measures provide information on the degree of abduction at which voicing ceased and started. In fully voiced tokens, these values are virtually identical to hFlowMax.
- (5) Automatic peak picking was performed in the AC flow signals for three pulses before and after voicing offset/onset in devoiced tokens, and for three pulses before/after the /h/ flow peak in voiced tokens. These three pulses were averaged to yield two single values, ACOff and ACon. These values reflect the amplitude of vocal fold vibration immediately adjacent to a voicing threshold (for devoiced tokens) or the /h/ flow maximum (for voiced tokens). Averaged values were used rather than single-pulse measures

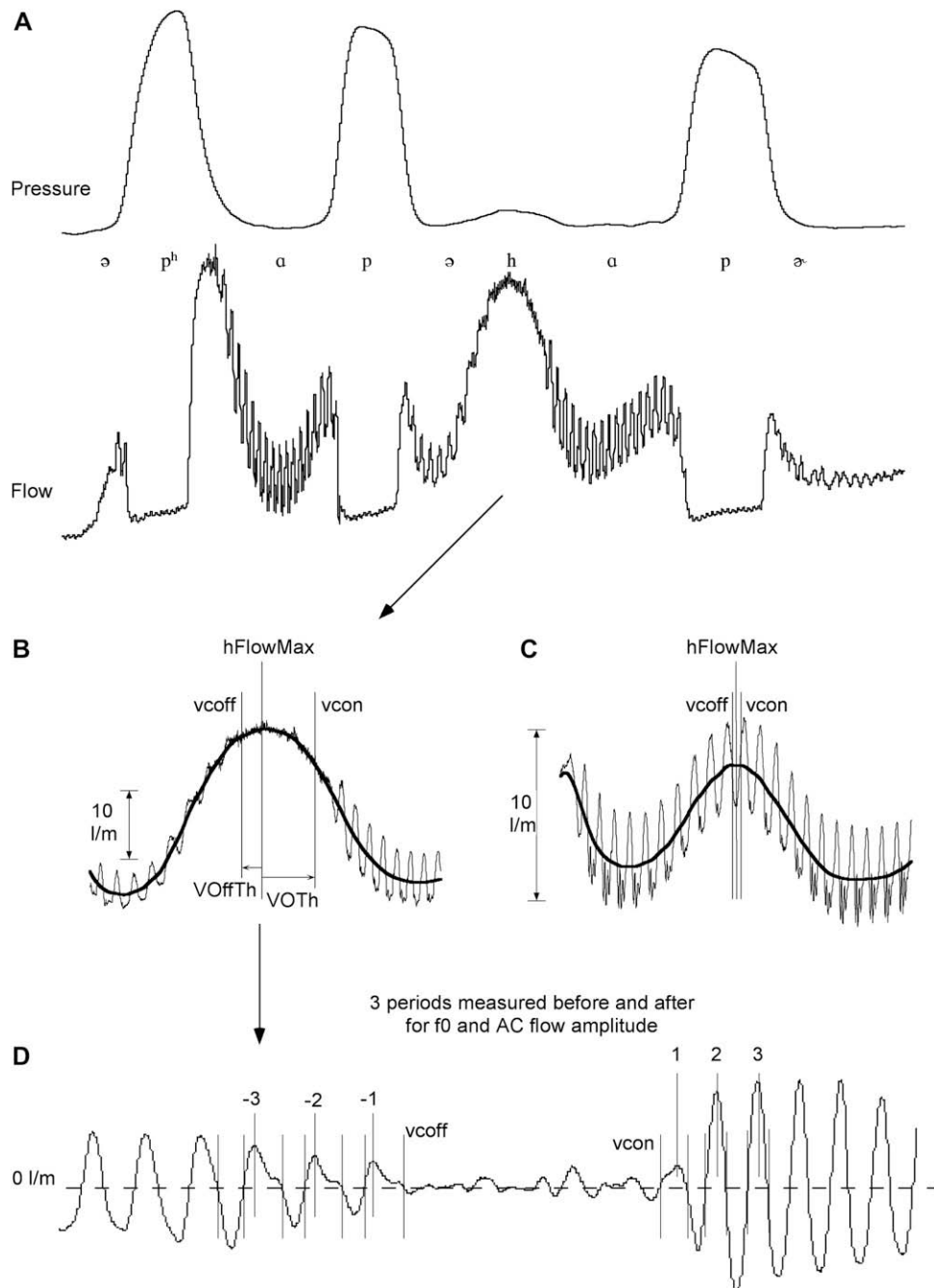


FIGURE 1. A. Flow and pressure signals for a full utterance; samples of (B) devoiced and (C) voiced productions of /h/, showing the labels for maximum flow in /h/ (hFlowMax), voicing offset (VcOff), and voicing onset (VcOn); and (D) an expanded section of the AC flow signal (original flow - DC flow) to show how the VcOff and VcOn times were established, as well as how zero-crossing and peak-picked labels were used to measure f0 and pulse amplitude. The same devoiced production is shown in panels A, B, and D. In cases of fully voiced /h/ (panel C), the VcOff and VcOn labels were set as close as possible to the hFlowMax label while allowing for a full number of glottal pulses before and after the hFlowMax label. The time duration of panel A is about 900 milliseconds. The duration of panels B and C is about 500 milliseconds. Notice that the vertical scale differs between panels B and C.

because we have found the former to have better test-retest reliability.¹⁴

(6) Pulse-by-pulse measures of f0 were obtained using zero-crossings in the AC flow signal. Analogous to the ACOff and ACOOn measures, three pulses were averaged before

and after voicing offset/onset or the /h/ flow maximum to yield f0Off and f0On values.

(7) To characterize f0 around /h/ in comparison to the surrounding vowels, f0 was also measured in the preceding, unstressed vowel and following, stressed vowel.

Single-pulse measures were obtained at the sixth glottal cycle preceding voicing offset and the tenth period after voicing onset. These periods were chosen so as to be some distance from the /h/ abduction, while providing values for at least 90% of the tokens. (For example, only 80% of the speaker's unstressed vowels had eight or more glottal cycles in block 2, so measuring f0 at eight cycles before the /h/ flow maximum would have involved not measuring 20% of the data.) For this speaker, six periods before voicing offset corresponded to a duration of about 50 milliseconds; 10 periods after voicing onset averaged about 68 milliseconds. These values are referred to as f0VowelPre and f0VowelPost.

- (8) To characterize our speaker's voice quality, we performed software inverse filtering for the utterance "Poppa Hopper" and measured open quotient (OQ) and speed quotient (SQ), after work suggesting that these aspects of pulse shape differentiate breathy, modal, and glottal fry phonation.²¹ (Inverse filtering, performed to minimize effects of the vocal tract transfer function, was not performed for the high vowels /i/ and /u/ because their low First Formants (F1s) may be close to the first harmonic (H1), and filtering may affect the H1 amplitude.) Filtering was performed over the regions between the negative- and positive-going zero-crossings in the original AC flow signal. This method approximates closed-phase filtering while still allowing semi-automatic processing. The resulting signal was smoothed twice iteratively with a five-point window to eliminate any discontinuities arising from the filtering. Smoothed (DC) and AC filtered flow signals were then obtained in the same manner as described above for the original flow signal. Finally, using this inverse-filtered AC flow signal, zero-crossing and peak-picking algorithms were applied and these times were used to determine the OQ and SQ for each pulse. OQ is the duration of the open phase divided by the period duration; SQ is the duration of the opening portion of the open phase divided by the closing portion. Our OQ measures obtained using zero-crossings essentially use a 50% criterion for defining open and closed phases, similar to what authors have suggested for OQ derived from electroglottographic signals.^{22,23} Dromey et al²² found that percentage-based measures of OQ are preferable to visually defined ones in being more reliable. We note that measures of OQ made using different signals and methods will differ in their absolute values; our main purpose here is not to present OQ measures to compare with other studies (there is a large body of normative OQ data in normal adult speakers), but rather to provide a reliable indication of how our speaker's phonatory behavior changed across the two recording blocks.

As with the f0 measures in the preceding and following vowels, the OQ and SQ measures were taken at the sixth pulse before voicing offset and the tenth pulse after voicing onset. Below, these values are indicated as OQVowelPre/Post and SQVowelPre/Post.

In sum, our measures provide information on several factors associated with phonation.

- (1) Pressure: subglottal driving pressure.
- (2) hFlowMax: maximum degree of abduction.
- (3) DCOff/On: the degree of abduction at voicing offset and onset (for voiceless tokens).
- (4) VOffTh and VOTH: the duration of the voicing break in /h/ (if any).
- (5) ACOff/On: the vibratory amplitude of the vocal folds in the vicinity of voicing offset/onset (or maximum flow in /h/, for fully voiced tokens).
- (6) f0Off/On and f0VowelPre/Post: variations in longitudinal tension of the vocal folds in the neighborhood of the abduction (f0Off, f0On) and in neighboring vowels (the f0Vowel measures).
- (7) OQVowelPre/Post and SQVowelPre/Post: laryngeal setting or phonation type (viz, degree to which the voice showed characteristics of breathiness: OQs and SQs close to 1).

RESULTS

Success in manipulating /h/ voicing

As indicated above, real-time observation during recording suggested that the /h/s in block 1 were mostly voiced. Subsequent analysis confirmed this: of 230 productions, only five were devoiced (2.2%). In block 2, when asked to devoice, our speaker produced 155/222 tokens with a voicing break (69.8%). This difference was highly significant ($\chi^2 = 225.03$, $P < 0.0001$). Thus, our speaker succeeded in changing his /h/ voicing patterns when asked.

Main effects for voicing versus devoicing condition

To determine which phonatory parameters changed across blocks, an analysis of variance (ANOVA) was performed, with independent variables of recording block (block 1 vs block 2), vowel [a i u], and loudness (S, N, L), and dependent variables VOffTh/VOTH, DCOff/On, hFlowMax, Pressure, ACOff/On, f0Off/On, and f0VowelPre/Post. Loudness and vowel were included as independent variables because we have observed¹⁴ that these factors may affect the likelihood of /h/ voicing, yet it was not clear whether any such effects would be identical across blocks. A separate ANOVA was performed on the OQ and SQ measures, with independent variables of block and loudness (recall that these measures were only performed for the vowel /a/). Table 1 shows the results of the ANOVAs. The table is broken into two parts to show the variables that were measured across all vowels (Table 1a) versus those that were made for the /a/ context only (Table 1b). Table 2 summarizes the directions of the block effects.

The significant block effects on VOffTh and VOTH (Table 1) simply indicate that in block 1 these values were approximately 0 milliseconds, whereas in block 2 they were usually positive

TABLE 1.
ANOVA Results: (a) Variables Measured Across All Vowels and (b) Variables Measured for /a/ Only

Variable(s)	Block	Loud	Vowel	Block × Loud	Block × Vowel	Loud × Vowel	Block × Loud × Vowel
(a)							
VOffTh	*	*		*			
VOTh	*	*		*			
DCOff	*	*	*	*	*	*	
DCOn	*	*	*	*	*	*	
hFlowMax	*	*	*	*	*	*	
Pressure	*	*	*	*		*	
f0Off	*	*	*	*			*
f0On	*	*	*	*			
f0VowelPre		*		*		*	
f0VowelPost	*	*	*				
ACOff	*	*	*	*		*	*
ACOn	*	*	*	*		*	*
DFs	1,413	2,413	2,413	2,413	2,413	2,413	4,413
(b)							
OQVowelPre	*	*	n/a		n/a	n/a	n/a
OQVowelPost	*	*	n/a		n/a	n/a	n/a
SQVowelPre	*	*	n/a	*	n/a	n/a	n/a
SQVowelPost		*	n/a	*	n/a	n/a	n/a
DFs	1,132	2,132	2,132	2,132	2,132	2,132	4,132

The three independent variables are recording block (block 1 vs block 2), loudness (S, N, L), and vowel (i, a, u). See text for description of the dependent measures. The last row in (a) and (b) provide the degrees of freedom (DFs) for the analyses. Asterisks indicate significance at $P < 0.01$. The "n/a" (not applicable) notation in the cells for OQ and SQ indicate conditions not relevant for these measures (made for the vowel /a/ only).

(c. 20–30 milliseconds; cf. Table 2), reflecting a voicing break. The block effect for hFlowMax shows that the speaker abducted to a greater degree in block 2. DCOff values were also significantly higher in block 2. DCOn values went in the same

direction, but the differences were not significant. This pattern of hFlowMax and DC flow results suggests that, in block 1, the speaker was not abducting enough to achieve devoicing. Table 2 shows, however, that the absolute block differences in hFlowMax and DCOff were quite small: only a few l/m. Subglottal pressure was lower in block 2 than in block 1. The f0 values were equivalent across blocks in the preceding vowel (f0VowelPre), but block 2 had lower f0s approaching the maximum flow in /h/ or voicing offset (f0Off) and higher f0s at voicing onset and in the following vowel (f0On, f0VowelPost). Block 2 also had lower glottal pulse amplitudes (ACOff/On) around the voicing break or hFlowMax, and higher OQ measures in both the preceding and following vowels. Finally, SQs were lower (closer to 1) in block 2 (significant only for the unstressed vowel).

Generally, these results indicate that our speaker made multiple sensible adjustments to achieve devoicing in block 2: (1) he increased his abduction degree; (2) he reduced his subglottal pressure; (3) he apparently increased the longitudinal tension of his vocal folds for the following stressed vowel, evident both at voicing onset and in the following vowel; and (4) he adjusted his vocal fold settings toward a breathy voice quality, evident in higher values of OQ (both preceding and following vowels) and an SQ closer to 1 (preceding vowel only). The lower ACOff/On values may reflect decreasing vibratory amplitudes that arise simply as a function of greater abduction degree, or may relate to an altered laryngeal setting (eg, differences in actively controlled tension).

TABLE 2.
Summary of Block Effects

Variable	Block 1 mean	Block 2 mean	Direction
VOffTh	4 ms	24 ms	Block 1 < Block 2
VOTh	5 ms	31 ms	Block 1 < Block 2
DCOff	13 l/m	15 l/m	Block 1 < Block 2
DCOn	13 l/m	14 l/m	Block 1 < Block 2 ^{NS}
hFlowMax	14 l/m	17 l/m	Block 1 < Block 2
Pressure	12 cm H ₂ O	11 cm H ₂ O	Block 1 > Block 2
f0Off	112 Hz	103 Hz	Block 1 > Block 2
f0On	123 Hz	135 Hz	Block 1 < Block 2
f0VowelPre	117 Hz	116 Hz	Block 1 < Block 2 ^{NS}
f0VowelPost	133 Hz	158 Hz	Block 1 < Block 2
ACOff	4 l/m	1 l/m	Block 1 > Block 2
ACOn	4 l/m	2 l/m	Block 1 > Block 2
OQVowelPre	0.41	0.45	Block 1 < Block 2
OQVowelPost	0.39	0.45	Block 1 < Block 1
SQVowelPre	1.38	1.17	Block 1 > Block 2
SQVowelPost	1.6	1.5	Block 1 > Block 2 ^{NS}

Results are significant unless otherwise noted.

^{NS} not significant. ms, milliseconds. Variable names are the same as in Table 1.

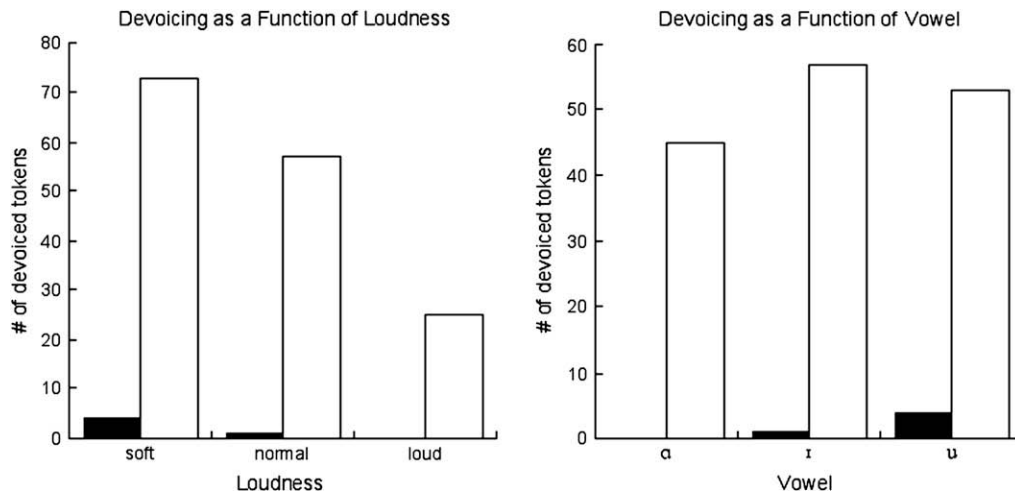


FIGURE 2. Effects of loudness (left) and vowel (right) on the number of devoiced tokens in each block. In each plot, the filled bars represent block 1 and the unfilled bars represent block 2.

Interactions among block, loudness, and vowel

Table 1 shows that vowel and loudness conditions had significant main effects on most dependent variables. More interesting for present purposes are the numerous block \times vowel or block \times loudness interactions. To clarify these interactions, we plotted the number of devoiced tokens in each block as a function of loudness and vowel. The results are shown in Figure 2. Because block 1 had very few devoiced tokens (represented by the filled bars), the main effects of vowel and loudness (cf. Table 1) reflect primarily the variations observed in block 2, and the interactions between vowel, loudness, and block presumably indicate that loudness and vowel mostly affected voicing when the speaker attempted to devoice. In block 2, the number of devoiced tokens varied with loudness in the order $S > N > L$, and devoicing was least frequent for the vowel /a/. Loudness

and vowel effects on the frequency of devoicing are consistent with what we have previously observed in women.¹⁴

Qualitative inspection of the data suggested a combination of respiratory, laryngeal, and supralaryngeal changes between blocks. Figures 3 and 4 illustrate some of these differences. Figure 3 shows the effects of loudness on DC flow for the two voicing blocks. In block 1, the effects of loudness on the hFlowMax (abduction) values are $L > S > N$; in block 2, they are $N > L > S$. The actual values of the maximum /h/ flow show that, going from block 1 to block 2, the speaker mainly increased abduction degree in the normal loudness condition (12 vs 20 l/m). The hFlowMax values increased slightly in the soft condition from block 1 to block 2 (14 vs 16 l/m), whereas values in the loud condition were virtually identical (c. 17 l/m in both blocks).

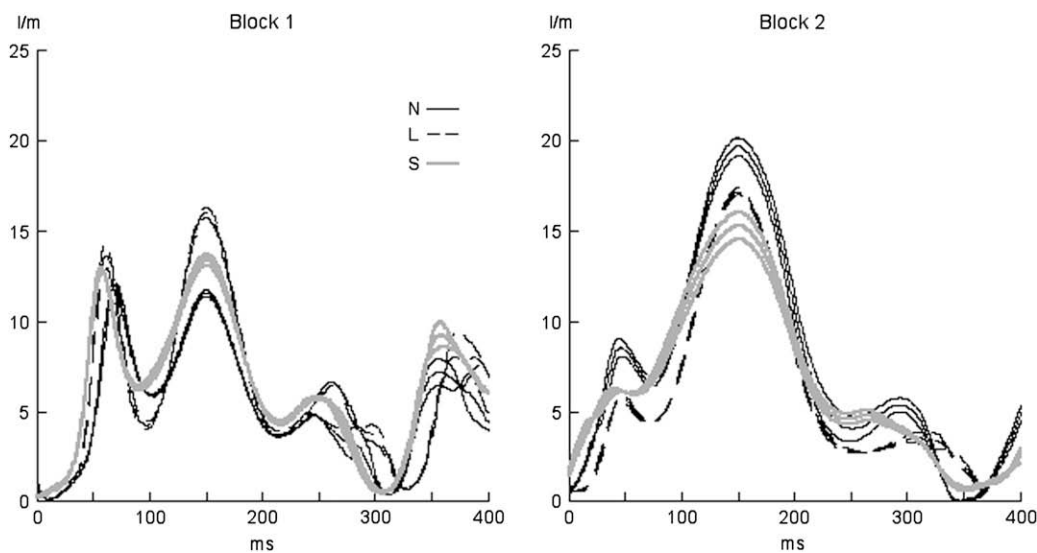


FIGURE 3. Loudness effects within each block: block 1 (left) and block 2 (right). In each plot, the plain solid lines represent the normal loudness condition (mean \pm 1 SD); the dashed lines represent the loud condition (mean \pm 1 SD); and the heavy gray lines represent the soft condition (mean \pm 1 SD).

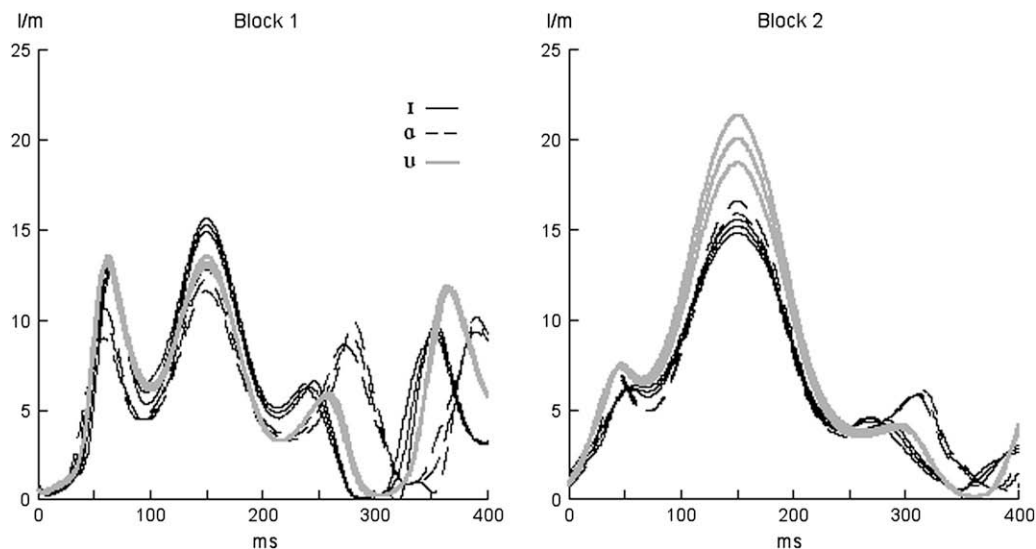


FIGURE 4. Vowel effects within each block: block 1 (left) and block 2 (right). In each plot, the plain solid lines indicated the /i/ context (mean \pm 1 SD); the dashed lines represent the /a/ context (mean \pm 1 SD); and the heavy gray lines indicate the /u/ context (mean \pm 1 SD).

Figure 4 shows the effects of vowel on DC flow across blocks. As with loudness, the vowel conditions do not have the same effects across blocks: from block 1 to block 2, peak DC values for /i/ remained essentially the same (c. 15 l/m); those for /a/ increased somewhat (12 vs 16 l/m); and those for /u/ increased most dramatically (13 vs 20 l/m).

Because loudness changes are achieved largely (though not entirely) by varying respiratory driving pressure, the changes shown in Figure 3 indicate that respiratory-laryngeal relationships differed across blocks. Because vowel variation mainly reflects supraglottal postures (which may in turn affect aspects of laryngeal position and tissue characteristics), the changes shown in Figure 4 indicate that laryngeal-supralaryngeal relationships differed across blocks.

Block effects independent of voicing differences

One question that arises is whether block differences are an artifact of voicing differences. That is, it might be that block comparisons made on tokens matched in voicing would not show significant differences. To test this, an ANOVA was performed on the voiced tokens only (225 from block 1; 67 from block 2). The dependent variables VOffTh, VOTh, DCOff, and DCOn were not included here because they are essentially irrelevant for fully voiced tokens: VOffTh and VOTh will be very close to zero, and DCOff/On will be very close to the maximum /h/ flow value. Further, because the assumption of homogeneity of variance was not met for this analysis, we used a more conservative significance criterion of $\alpha = .001$. The results are given in Table 3. For simplicity, only main effects and interactions involving the block variable are shown. Asterisks indicate effects that remain significant in this reduced analysis; NS indicates formerly significant effects that now fail to reach significance; empty cells indicate comparisons that were not significant in either analysis. Table 3 shows that, although some formerly significant relationships fall below significance

in this smaller data set, many (about half) of them do not. We conclude that the block effects are not an artifact of the number of voiceless tokens, but reflect general production differences across blocks. The loss of all significant effects for hFlowMax does suggest that greater abduction degrees in block 2 were mainly associated with tokens that were, in fact, devoiced.

TABLE 3.
ANOVA Results on Fully Voiced Tokens: Variables Measured for (a) All Vowels and (b) /a/ Context Only

Variable(s)	Block	Block \times Loud	Block \times Vowel	Block \times Loud \times Vowel
(a)				
hFlowMax	NS	NS	NS	
Pressure	NS	*		
f0Off	NS	*		NS
f0On	NS	*		
f0VowelPre		NS		
f0VowelPost	*			
ACOff	*	*		NS
ACOn	*	*		NS
DFs	1,263	2,263	2,263	3,263
(b)				
OQVowelPre	*		n/a	n/a
OQVowelPost	NS		n/a	n/a
SQVowelPre	NS	*	n/a	n/a
SQVowelPost		NS	n/a	n/a
DFs	1,99	2,99	2,99	3,99

Main effects of vowel and loudness are not shown because they are irrelevant to this comparison. The last rows in (a) and (b) provide the degrees of freedom for the analyses. Asterisks indicate significance at $P < 0.001$. "NS" indicates effects that were significant in the full analysis (Table 1) but failed to reach significance in this reduced analysis; the vowel interactions cells for OQ and SQ show "n/a" because these measures were restricted to the /a/ context.

TABLE 4.
Summary of Correlations Run Separately on the Two Blocks for All Variables Except the Voice Timing Variables VOTH and VOffTh (Excluded Because of Extreme Non-Normality)

	/a/	/l/	Loud	DCOff	hFlowMax	DCOn	Pressure	f0Off	f0On	f0VowelPre	f0VowelPost	ACOff	ACOn
/a/-ness	1.00												
/l/-ness	-/-	1.00											
Loudness			1.00										
DCOff	-/NS	+/-	+/+	1.00									
hFlowMax	-/NS	+/-	+/NS	+/+	1.00								
DCOn	-/NS	+/-	+/+	+/+	+/+	1.00							
Pressure		-/-	+/+	+/+	+/+	+/+	1.00						
f0Off			+/+	+/NS	+/NS	+/NS	+/+	1.00					
f0On	NS/-	NS/+	+/+	+/NS	+/NS	+/+	+/+	+/+	1.00				
f0VowelPre			+/+	+/+	+/NS	+/+	+/+	+/+	+/+	1.00			
f0VowelPost			+/+	+/+	+/NS	+/+	+/+	+/+	+/+	+/+	1.00		
ACOff	-/NS		+/+	+/+	+/+	+/+	+/+	+/+	+/NS	+/+	+/+	1.00	
ACOn	-/NS		+/NS	+/+	+/+	+/+	+/+	+/NS	+/NS	+/NS	+/NS	+/+	1.00

For simplicity, only the directions of significant correlations are shown; empty cells indicate that the *r* value was not significant for either block; NS, not significant at *P*<0.01. Results for block 1 are on the left of the slash, and block 2 are on the right. Outlined cells indicate that the transformed *r* values differed significantly between the two blocks at *P*<0.01.

Interrelationships among variables

The analyses above suggest that our speaker altered multiple production parameters when attempting to devoice, and that the relationships among laryngeal, supralaryngeal, and respiratory variables also changed. In this section, we investigate these interrelationships more formally, using correlations and principal components analysis (PCA). For these analyses, vowel was recoded into two quasi-continuous variables of “/a/-ness” (/a/ = 1; /u, ɪ/ = 0) and “/l/-ness” (/l/ = 1; /a, u/ = 0). The voicing variables VOTH and VOffTh were not included because the very small number of devoiced tokens (five of 230) in block 1 meant that most values were close to zero, and correlations would reflect only the characteristics of a small set of the data. Thus, these analyses do not directly speak to how the speaker achieved voicing or devoicing *per se*, but rather show general production differences between blocks. Also, because this analysis combines measures from preceding and following vowels, and voicing offsets and onsets, some of the results indicate rather global aspects of the utterances. Finally, the voice quality measures OQ and SQ are not included here because they were only made for the /a/ context.

Table 4 summarizes the correlations for the two blocks. The results are based on the 231 tokens (216 from block 1, 215 from block 2) where all dependent measures were available. For simplicity, only the directions (+ or -) of significant (*P* < 0.01) correlations are shown, with results for both blocks in each cell (block 1/block 2). Correlations that were significantly different (*P* < 0.01) between the two blocks after an *r*-to-*z'* transformation are indicated in boldface. (The *r*-*z'* transform corrects for nonnormality of the *r* distribution.)

The data show that many correlations are significant in both blocks, again probably reflecting general production settings. For example, in both analyses, f0 values in preceding and following vowels are correlated. Also, strong (*r* > 0.89) correlations were observed across blocks between hFlowMax and

DCOff as well as DCOn. Given that vocal fold vibration shows a hysteresis effect,^{24–26} even tokens with voicing breaks frequently show voicing offset very close to the time of the maximum /h/ flow peak. Our past work¹⁴ has also indicated that DC flow values at voicing onset are usually significantly correlated with hFlowMax. Thus, it is not surprising that these relationships persist across the two input conditions for our speaker.

Despite such commonalities, however, there are also many differences across blocks. For example, significant effects of vowel on DCOff, hFlowMax, and DCOn in block 1 disappear in block 2, as do relations between hFlowMax and the four f0 variables (f0Off/On, f0VowelPre/Post). Conversely, vowel variation is associated with changes in f0On in block 2 but not block 1. Thus, the correlational results support the claim that interrelationships among laryngeal, respiratory, and supralaryngeal variables changed when the speaker attempted to devoice his /h/s.

As a final investigation into production changes between blocks, we performed PCA on all variables except VOTH and VOffTh (excluded for the same reasons noted above for the correlations), and compared the resulting factor structures across blocks. PCA reduces the dimensionality of the data and shows which variables are most strongly correlated. Highly correlated variables load together on a single factor, whereas correlations are minimized across factors. Variables were retained using the larger of two values based on changes in the scree plot and a 75% variance criterion.

The orthogonal factor structures (rotated using the Varimax procedure) are shown in Table 5. The solutions for both blocks showed significant (*P* < 0.0001) χ^2 values, indicating that they were highly effective in characterizing the data. From 16 input variables, three factors were extracted for block 1, and four factors were extracted for block 2.

As with the correlation results, there are some consistencies across blocks. Both have a factor with heavy loadings of

TABLE 5.
Orthogonal Factor Solutions from the PCA

	Block 1			Block 2			
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 4
/a/-ness			-0.790			-0.843	
/i/-ness			0.848			0.852	
Loudness	0.907				0.741		0.559
DCOff		0.966		0.943			
hFlowMax		0.970		0.957			
DCOn		0.965		0.898			
Pressure			-0.445				0.829
f0Off	0.976				0.886		
f0On	0.925				0.703		
f0VowelPre	0.976				0.914		
f0VowelPost	0.955				0.933		
ACOff	0.877			0.730			
ACOn	0.891			0.786			

To simplify the display, only factor loadings of 0.4 or greater are shown.

DCOff, hFlowMax, and DCOn (factor 2 in block 1 and factor 1 in block 2). As indicated above, these variables are highly correlated in all speakers we have studied, so it is not surprising that these relationships are maintained across blocks. Also, the four f0 variables load together on a single factor in both blocks, presumably indicating that all of these reflect, to some degree, the overall f0 setting for the utterance. In other respects, though, factor structures differ across blocks. For example, in block 1, subglottal pressure varies with vowel quality, but in block 2, it is more closely related to loudness condition. The loudness manipulation covaries with pulse amplitude (ACOff, On) and with the four f0 variables in block 1, but in block 2, loudness covaries with the f0 variables only, with AC flow amplitudes being more closely related to the DC flow variables (DCOff, DCOn, and hFlowMax). Thus, the PCA results are consistent with the correlations in suggesting that not only did the speaker make system-wide alterations across blocks, but interrelationships among respiratory, laryngeal, and supra-laryngeal variables also changed.

DISCUSSION

Management of voicing

Our speaker was quite successful in devoicing his /h/ productions on request. He achieved this via a combination of greater abduction, lower subglottal pressures, and greater longitudinal tension of the vocal folds for the stressed vowel (reflected in higher f0 values immediately following /h/ and in the following vowel). All of these factors have been associated with higher phonation threshold pressures in more formal experiments.^{8,11,25} Voice-source measures suggested a breathier voice quality in block 2, with higher OQ values (both preceding and following vowels) and SQs closer to 1 (preceding vowel only). Finally, glottal pulse amplitudes around /h/ were lower when devoicing was achieved than when it was not. This could reflect simply the inhibitory effect of greater abduction on vocal fold

vibration, and/or more general effects of laryngeal setting (eg, degree of longitudinal tension).

Whereas the physiological underpinnings of Pressure (ie, respiratory driving pressure) and hFlowMax (ie, abduction degree) are fairly straightforward, the f0 and voice quality results require a bit more discussion.

Laryngeal setting: f0, OQ, and SQ

Fundamental frequency is affected by active (muscularly controlled) changes in longitudinal tension, passive stretching during abduction, and variation in subglottal pressure.²⁷ The data in Table 2 suggest that the two blocks were characterized by actively controlled differences in longitudinal tension in the /h/-initiated syllable. The absolute f0 differences across blocks were 25 Hz in the middle of the stressed vowel (133 vs 158 Hz), but only 12 Hz at voicing onset (123 vs 135 Hz). Higher abduction degrees in block 2 could have two opposing effects on f0 in the immediate vicinity of /h/: greater passive stretching should increase f0, whereas a greater drop in subglottal pressure caused by decreased laryngeal resistance should lower f0. The smaller block difference at voicing onset compared to the stressed vowel suggests that a lowered subglottal pressure during abduction in block 2 mitigated f0 increases related to both active and passive tension changes. In the middle of the vowel, this effect would have dissipated, so that f0 values would reflect mostly differences in muscle use patterns.

The OQ reflects the degree and duration of glottal closure in the voicing cycle; higher OQs are associated with incomplete glottal closure and breathier voice qualities. In simple models of vocal fold biomechanics such as the two-mass model we have used in the past,¹² an OQ of 1 indicates that vocal fold oscillation amplitude is less than or equal to the resting glottal half-width,⁸ such that the vocal folds do not come together in a measurable closed phase. In these models, OQ decreases from 1 as oscillation amplitude (reflected in our data AC flow values) increases relative to glottal width (reflected in our

data as DC flow values). The Q-factor (a scaling factor for the natural frequencies of the model) also affects OQ in simulations: as Q (and thus f_0) increases, the DC flow decreases but reaches a plateau, whereas the AC flow declines continuously, altering the AC-DC flow ratio and thus OQ. In sum, the increased OQs in block 2 appear to reflect a more abducted glottal posture, possibly with contributions from increased vocal fold tension.

Increases in SQ have been associated with greater vocal tract loading.²⁸ According to Titze's load quotient²⁹ based on a parametric description of the glottal source function, skew (ie, SQ) should decrease with smaller oscillation amplitudes, larger sub- and supraglottal areas, and higher subglottal pressure. (The load quotient equation also includes an empirically defined coefficient k , but it is unclear how speakers could manipulate this directly, so we focus on the parameters that have clear physiological correlates.) As indicated above, subglottal pressures were generally lower in block 2, so this cannot be the explanation for lower SQ values. Further, given that vowel quality was controlled across blocks, it is unlikely that differences in sub- and supraglottal areas account for differences in SQ. Thus, the most sensible explanation for lower SQ values in block 2 is reduced oscillation amplitudes, reflected in lower AC flow values. Although AC flow was measured adjacent to the voicing offset/onset or hFlowMax, SQ was measured at some distance from the /h/. This suggests that the speaker's glottal setting differed across the entire VCV sequence, not only in the vicinity of /h/.

Interrelationships among variables

Several aspects of the results indicate that our speaker made system-wide changes between blocks. Differing vowel and loudness effects on DC flow contours (Figures 3 and 4) across blocks suggest altered relationships among supralaryngeal, respiratory, and laryngeal systems. These differences were not limited to a narrow window around /h/, but affected the adjacent vowels and the subglottal pressures measured in the flanking /p/ closures. These block differences are consistent with the interactions observed between block, loudness, and vowel (Tables 1 and 2). Many block effects persisted when the ANOVA was run on fully voiced tokens only (Table 3), indicating that the differences were not an artifact of whether the speaker devoiced or not. Correlations showed that although some strong, typical relationships among phonatory variables remained consistent across blocks, many relationships changed. The loading patterns in the PCA further support the conclusion that phonatory variables correlated in different ways across the two blocks.

Implications

The factors that influence phonation threshold pressure have been established in formal mathematical and modeling studies, but few data are available on how individual speakers vary these parameters to achieve specific phonatory goals. This experiment represents a naturalistic exploration of how a single speaker manipulated the likelihood of phonation. The varied and complex pattern of differences between the two recording blocks suggests that our subject had, at some level, a rather

sophisticated awareness of the factors involved in voicing control. He did not adjust a single parameter (such as increased abduction degree) to achieve devoicing, but rather made system-wide changes in multiple, appropriate ways.

It is true that our speaker had a history of medical intervention for voice-related issues, and his surgery and current medication may have affected some aspects of his phonatory characteristics. It is hard to imagine, however, how these factors, on their own, could lead to the general findings obtained here: namely that he was able to alter his voicing behavior on request, and that he made large-scale changes in his speech production settings to do so. It could be that our speaker's level of success in changing phonatory behavior partly reflected his phonetic training. Our cross-speaker work does suggest that individuals may differ widely in precisely how they manage voicing, and it is possible that some speakers would use a more limited range of strategies.

CONCLUSIONS

The current results indicate that at least some (phonetically trained) speakers can make major changes in their voicing behavior around an abduction gesture, and that they may adjust a wide range of physiological parameters to do so. The patterns of changes we found are consistent with our expectations based on formal studies of phonation threshold pressure, but this experiment differs from past work in being naturalistic, that is, investigating how a living speaker controls phonation in running speech. Our subject showed a tacit understanding of the full range of variables that affect phonation thresholds.

Acknowledgments

We are grateful to our speaker for his participation and for providing us with detailed information on his vocal and professional history. We also thank Anders Löfqvist and two anonymous reviewers for comments on earlier versions of this manuscript. This work was supported by NIH grant DC-00865 to Haskins Laboratories and by CNPq, Brazil.

REFERENCES

1. Jones D. *The Pronunciation of English*. 4th ed. London: Cambridge University Press; 1956.
2. Abercrombie D. *Elements of General Phonetics*. Chicago, IL: Aldine Press; 1967.
3. Ladefoged P. *A Course in Phonetics*. 3rd ed. Orlando, FL: Harcourt Brace College Publishers; 1993.
4. Ishizaka K, Flanagan JL. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst Tech J*. 1972;51:1233-1268.
5. Stevens KN. Physics of laryngeal behavior and larynx modes. *Phonetica*. 1977;34:264-279.
6. Titze IR. The physics of small-amplitude oscillation of the vocal folds. *J Acoust Soc Am*. 1988;83:1536-1552.
7. Titze IR. Physiologic and acoustic differences between male and female voices. *J Acoust Soc Am*. 1989;85:1699-1707.
8. Titze IR. Phonation threshold pressure: a missing link in glottal aerodynamics. *J Acoust Soc Am*. 1992;91:2926-2935.
9. van den Berg J. Myoelastic-aerodynamic theory of voice production. *J Speech Hear Res*. 1958;1:227-244.
10. Koenig LL. Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *J Speech Lang Hear Res*. 2000;43:1211-1228.

11. Chan RW, Titze IR, Titze MR. Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa. *J Acoust Soc Am.* 1997;101:3722-3727.
12. Lucero JC, Koenig LL. Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control. *J Acoust Soc Am.* 2005;117:1362-1372.
13. Lucero JC, Koenig LL. Phonation threshold pressures as a function of laryngeal size in a two-mass model of the vocal folds. *J Acoust Soc Am.* 2005;118:2798-2801.
14. Koenig LL, Mencl WE, Lucero JC. Multidimensional analyses of voicing offsets and onsets in female speakers. *J Acoust Soc Am.* 2005;118:2535-2550.
15. Whalen DH, Levitt AG. The universality of intrinsic f_0 of vowels. *J Phon.* 1995;23:349-366.
16. Orlikoff RF. Vocal stability and vocal tract configuration: an acoustic and electroglottographic investigation. *J Voice.* 1995;9:173-181.
17. Löfqvist A, Carlborg B, Kitzing P. Initial validation of an indirect measure of subglottal pressure during vowels. *J Acoust Soc Am.* 1982;72:633-635.
18. Smitheran JR, Hixon TJ. A clinical method for estimating laryngeal airway resistance during vowel production. *J Speech Hear Disord.* 1981;46:138-146.
19. Stevens KN. *Acoustic Phonetics.* Cambridge, MA: MIT Press; 1999.
20. Löfqvist A. Acoustic and aerodynamic effects of interarticulator timing in voiceless consonants. *Lang Speech.* 1992;35:15-28.
21. Childers DG, Lee CK. Vocal quality factors: analysis, synthesis, and perception. *J Acoust Soc Am.* 1991;90:2394-2410.
22. Dromey C, Stathopoulos ET, Sapienza CM. Glottal airflow and electroglottographic measures of vocal function at multiple intensities. *J Voice.* 1992;6:44-54.
23. Rothenberg M, Mahshie JJ. Monitoring vocal fold abduction through vocal fold contact area. *J Speech Hear Res.* 1988;31:338-351.
24. Berry D, Herzel H, Titze IR, Story B. Bifurcations in excised larynx experiments. *NCVS Status Prog Rep.* 1995;8:15-24.
25. Lucero JC. The minimum lung pressure to sustain vocal fold oscillation. *J Acoust Soc Am.* 1995;98:779-784.
26. Lucero JC. A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset. *J Acoust Soc Am.* 1999;105:423-431.
27. Titze IR. *Principles of Voice Production.* Englewood Cliffs, NJ: Prentice-Hall; 1994.
28. Rothenberg M. Acoustic interaction between the glottal source and the vocal tract. In: Stevens KN, Hirano M, eds. *Vocal Fold Physiology.* Tokyo: University of Tokyo Press; 1981:305-323.
29. Titze IR. Parametrization of the glottal area, glottal flow, and vocal fold contact area. *J Acoust Soc Am.* 1984;75:570-580.