Original Paper

Folia Phoniatrica et Logopaedica

Folia Phoniatr Logop 2002;54:223–239
DOI: 10.1159/000065199

Segmental Level Analysis of Laryngeal Function in Persons with Motor Speech Disorders

Kate Bunton[a] Gary Weismer[b]

[a]National Center for Neurogenic Communication Disorders, University of Arizona, Tucson, Ariz.; [b]Waisman Center and Department of Communication Disorders, University of Wisconsin, Madison, Wisc., USA

Key Words
Speech intelligibility · Motor speech disorders · Speech acoustics · Laryngeal function · Speech perception

Abstract
Laryngeal behavior for segmental function is often disturbed in motor speech disorders. Loss of voicing contrasts has been shown to significantly contribute to speech intelligibility deficits. The present study was designed to examine two commonly erred laryngeal contrasts, the word-initial voiced-voiceless and glottal-null contrasts using acoustic analysis techniques. Acoustic measures were compared to expectations for the contrast based on data in the literature as well as listeners' perception of the token. Findings indicate a mismatch between acoustic data and both expectations for the contrasts and listener perception. There is some indication that changes in laryngeal segmental function are related to aging in general and may be exaggerated in persons with motor speech disorders.

Introduction

One of the most common characteristics of dysarthria is the impairment of phonatory function. Monopitch, monoloudness, and harshness are the most frequently mentioned dimensions of abnormality, across several dysarthria subtypes [1]. A number of studies have reported a relationship between these phonatory disruptions and speech intelligibility [2]. The suprasegmental variables and laryngeal articulatory variables of pitch, prosody, and the voicing distinction have been studied in individuals with hearing impairment, alaryngeal voice and motor speech disorders, and have been related to reductions in

KARGER
Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

Accessible online at:
www.karger.com/journals/fpl

Kate Bunton, PhD
Institute for Neurogenic Communication Disorders
University of Arizona, PO Box 210071
Tucson, AZ 85721 (USA), Tel. +1 520 621 1472
Fax +1 520 621 2226, E-Mail bunton@u.arizona.edu

intelligibility in those populations [3, 4]. However, a review of the literature suggests that the role of phonatory characteristics in intelligible speech production is relatively limited when compared to segmental contributions [5–10]. One study done by Kent et al. [11] examined laryngeal function from two perspectives: segmental (phonetic) function, expressed as features in an analytic intelligibility test, and commonly used acoustic measures of sustained vowel phonation (often taken as a primary task for evaluation of phonatory quality). Results showed there were differences between groups in segmental laryngeal function, but not in measures typically associated with vocal function, such as the acoustic measures of F_0, jitter, shimmer, and S/N ratio. Segmental analysis, therefore, may be more successful in establishing the acoustic correlates of laryngeal function and their contribution to speech intelligibility deficits.

The voiced-voiceless contrast is the primary articulatory function of the larynx that has been studied in relation to speech intelligibility [8, 12]. More recently, the initial glottal versus null (e.g., /hat/ versus /at/) contrast has been considered [13, 14]. There are only two studies that suggest the voicing distinction may not be produced clearly by dysarthric speakers [15, 16], however, numerous perceptual studies have found voicing errors to be common in the speech of persons with parkinsonian, spastic, and athetoid dysarthria [11, 14, 17–19]. Presence of errors involving /h/ versus initial vowel is supported on the basis of perceptual data by reports of laryngeal impairment across etiology and gender [14, 17, 20–23]. However, there is no clear indication that voicing errors are typical of dysarthric populations exclusively. Aging in general seems to be associated with a partial loss of voicing control, at least in the syllable-initial position [21, 24]. It also appears that there is an interaction between neurologic disease and gender for this aspect of voice function [22, 23].

The Kent et al. [14] intelligibility test, which was designed to allow a systematic investigation of perceptual and acoustic correlates of the speech signal to examine sources of reduced intelligibility in word level tests, targets three phonetic contrasts which relate to laryngeal segmental function: syllable-initial voicing contrast, syllable-final voicing contrast, and syllable-initial /h/-vowel contrast. This intelligibility test allows for interpretation of confusions in the listener's responses with respect to the features used in construction the list of test words, thus allowing examiners to create a relatively complete error profile. Previous reports of error profiles constructed using the Kent et al. [14] intelligibility test have shown that features involving laryngeal function are more commonly involved than other segmental features [11, 14, 17].

For speakers in the present study, phonetic contrast errors associated with laryngeal function were among the top 5 errors across speaker groups. Table 1 shows the top 8 errors for each speaker group in decreasing order of magnitude. It appears that segmental functions of the larynx are particularly vulnerable in word-initial positions [see also ref. 25]. Although the voiced-voiceless distinction in word-final position was among the top errors for three of the seven speaker groups, the frequency of the errors was significantly lower than the other two laryngeal contrasts [26]. For the word-initial voicing contrast, word-initial /h/ vs. vowel contrast, and word-final voicing contrast the error rates were 16, 29, and 9%, respectively.

The present study assumed that phonetic contrast errors derived from the intelligibility test would have well-defined acoustic underpinnings, as derived from research on acoustic phonetics and the perceptual cues for

Table 1. The most frequently occurring phonetic contrast errors for each speaker group listed in descending order

Geriatric female	Geriatric male	ALS female	ALS male	PD female	PD male	CVA male
high-low vowels	high-low vowels	stop-nasal	*glottal-null*	high-low vowels	*glottal-null*	*glottal-null*
Voiced-voiceless initial consonant	*glottal-null*	high-low vowels	final consonant-null	stop-fricative	high-low vowels	high-low vowels
Stop-nasal	Long-short vowels	voiced-voiceless final consonant	high-low vowels	stop-affricate	*voiced-voiceless initial consonant*	*voiced-voiceless initial consonant*
Other fricative place of articulation	stop-nasal	stop place of articulation	*voiced-voiceless initial consonant*	*glottal-null*	alveolar palatal fricative	alveolar palatal fricative
Final cluster-singleton		fricative-affricate	voiced-voiceless final consonant	*voiced-voiceless initial consonant*	voiced-voiceless final consonant	stop-nasal
		final consonant-null	/r/-/w/	/r/-/w/	initial consonant-null	stop-fricative
		initial cluster-singleton	stop-nasal	alveolar palatal fricative	initial cluster-singleton	/r/-/w/
		stop-fricative	initial cluster-singleton	stop-nasal	stop-fricative	stop-affricate

The two contrasts italicized in the table were selected for acoustic analysis in the present study.

sound segments [14]. The strategy of identifying segmental-level cues in syllable or word-level units has a long history in the speech perception literature [see ref. 27], but has been applied only sparingly in world identification error patterns produced by persons with speech disorders. In previous work looking at vowel tongue-height contrasts, however, an apparent mismatch between acoustics and perception was found [26, 28]. Despite this finding, sustained research efforts are required to develop an acoustically based, predictive model of speech intelligibility deficits in dysarthria. These investigations should examine further the relationship between an articulatory deficit and the perceptual phenomena that contribute to speech intelligibility across contexts. In a continuing effort to understand the acoustic basis of intelligibility deficits in dysarthria, the present study was designed to examine the question of whether there was a straightforward relationship between underlying acoustic cues for laryngeal contrasts (voice-voiceless and glottal-null) and perception of these contrasts for persons with motor speech disorders.

Methods

Participants

Four groups of speakers were included in the present study, including: (1) 10 speakers who had no history of neurologic disease (age-matched controls, normal geriatric, NG), (2) 10 speakers diagnosed as having amyotrophic lateral sclerosis (ALS), (3) 10 diagnosed with Parkinson's disease (PD), and (4) 5 speakers with a history of cerebrovascular accident (CVA). All speakers selected for the current study were recruited as part of a larger study on dysarthria. The speakers within each group were diverse in several respects, including severity of dysarthria, duration of

Table 2. Individual speaker characteristics

Speaker group	Speaker	Age years	Gender	Duration years	Lesion location	Intelligibility, %	
						mean	SD
NG	1	68	M			94.38	0.67
NG	2	69	M			95.24	2.79
NG	3	72	M			95.52	0.60
NG	4	68	M			97.96	0.45
NG	5	77	M			94.52	2.28
NG	6	68	F			96.67	1.38
NG	7	72	F			96.38	1.21
NG	8	76	F			98.38	0.61
NG	9	76	F			97.96	1.21
NG	10	68	F			97.38	1.05
ALS	11	61	M	1		85.4	4.27
ALS	12	41	M	2		78.5	6.45
ALS	13	62	M	12		83.4	1.66
ALS	14	55	M	1		92.7	2.76
ALS	15	29	M	2		93.8	2.87
ALS	16	57	F	1		75.65	3.80
ALS	17	47	F	–		76.5	6.82
ALS	18	75	F	1		91.23	3.60
ALS	19	52	F	9		96.24	2.62
ALS	20	52	F	10		91.52	2.80
PD	21	62	M	3		90.09	3.93
PD	22	64	M	8		82.94	3.81
PD	23	73	M	7		85.51	5.08
PD	24	62	M	9		86.51	3.83
PD	25	81	M	13		91.38	3.19
PD	26	72	F	20		89.52	3.76
PD	27	79	F	6		90.95	3.58
PD	28	67	F	4		94.67	2.68
PD	29	82	F	1		92.66	4.09
PD	30	75	F	17		94.81	3.02
CVA	31	71	M	1	left	79.6	7.11
CVA	32	77	M	1	left	92.9	3.24
CVA	33	58	M	1	right	86.7	5.23
CVA	34	75	M	3	right	77.1	6.51
CVA	35	63	M	1	bilateral	89.7	4.05

Blanks within the chart indicate that heading was not applicable. Speaker groups include NG, ALS, PD, and CVA. Duration refers to the time since diagnosis. Intelligibility scores are based on the Kent et al. [14] test.

the disease, medical history, and social substance abuse. A lower cutoff of 75% on the Kent et al. [14] single-word intelligibility test was used when selecting subjects, as meaningful acoustic analysis is more difficult with less intelligible speech. Individual speaker characteristics can be found in table 2. The three disorder groups chosen in the present study, ALS, PD, and CVA, were selected largely because of the availability of subjects in a referral pool and because the literature indicates that speech characteristics may vary with disease pathophysiology.

Intelligibility Measures

Intelligibility data and error profiles were obtained from a word-identification test with a multiple-choice format designed to examine 19 phonetic errors common in persons with motor speech disorders [14]. Error rates for each of the 19 contrasts were calculated for each speaker by recording each time a listener marked a response other than the target word, and dividing the total errors for each word pair by the number of listeners (n = 10). For instance, if the target was *had* and 3 of the listeners chose *add*, an error rate of 0.3 (3/10) would be marked in the glottal-null feature category. For the present study tokens were considered errors if 2 or more listeners marked a response other than the target (error rate of 0.2 or greater). Detailed recording and scoring procedures can be found in Kent et al. [14, 29].

Acoustic Measures

Voiced-Voiceless Contrast

Several acoustic correlates of the word-initial voicing contrast have been proposed including voice onset time (VOT), presence of voicing in the closure interval, and possibly the duration of the adjacent vowel [30–33]. For voiced stops, periodic energy resulting from vocal fold vibration that either preceded or was nearly coincidental with release was expected, whereas for voiceless stops, voice onset typically follows release by 40–100 ms [32, 34]. During the delay between release and voice onset, there should be little or no energy in the region of F_1 (the so-called F_1 cutback). A natural covariance between the duration of F_1 cutback and F_1 onset frequency, reflected in the response patterns to a voiced-voiceless continuum, has been presented in the perceptual literature [35]. No comprehensive investigations of the various acoustic counterparts of the voicing distinction have been done in dysarthric speech, so although acoustic correlates for this contrast have been defined in 'normals' it is unclear whether they will capture the distinction effectively in the dysarthric population. Further, there is also evidence that the voicing status of stops can be cued by a large number of acoustic dimensions, and redundancy of the cues may explain the perceptual resiliency of the contrast as reported by Miller and Nicely [36]. The acoustic measures selected for the present study are, therefore, a best guess at the physical basis of the perceptually based phonetic contrast for voicing of word-initial stops.

Measurements of the target words for this contrast included: (1) the VOT, (2) the presence or absence and (if present) duration of voicing preceding the burst [30–33], (3) frequency of F_1 at the onset of voicing and (4) vowel duration measured from the first glottal pulse to the last. Initial fricative duration was also measured for one of the word pairs *(sip-zip)*, since this has been identified as a strong cue for fricative voicing [37].

All measurements were made using a combined spectrogram/waveform display in CSpeech [38]. Nine possible word pairs were analyzed for this contrast: *pat-bat, bad-pad, pit-bit, sip-zip, coat-goat, dug-tug, cash-gash, tile-dial,* and *bunch-punch*. Target words are listed first for each pair. Tokens referred to as errors throughout this paper reflect the perception of the token based on the responses of the 10 listeners, and not necessarily errors in production. Of the total of 315 tokens (9 tokens × 35 speakers) to be measured, 47 were considered errors.

Initial /h/-Vowel

Clearly outlined acoustic descriptors of /h/ are scarce in the literature, but available reports suggest the acoustic characteristics of the /h/ are context-dependent [39, 40]. An /h/ is considered to be a voiceless glottal fricative, however, it is often phonetically voiced. Whether or not /h/ is voiced is determined in part by phonetic context. The /h/ voicing quality is influenced by the voicing feature of the adjacent sound segments such that in a voiced phonetic environment (e.g., in the word *Ohio*) it can become phonetically voiced. The targets in the Kent et al. [14] intelligibility test were all word-initial and should remain voiceless, but because the adjacent vowels are voiced the /h/ may take on the voicing feature [39]. Characteristics of initial /h/ have also been described by Lehiste [40]. She reported that an initial /h/ has a formant structure similar to that of the following syllable nucleus, however, the first formant is typically absent and the frequency positions for F_2-F_3 are slightly higher than those of the following vowel [40]. The presence of a split formant in the region of F_2 was also found to be characteristic of /h/ noise. The two formant-like bands of energy in the region of the second formant merge into the F_2 of the following vowel [40]. One additional segmental acoustic measure that has been reported to correlate with production of an /h/ is the presence of a periodic high-frequency noise energy (aspiration) [33, 34, 41].

Measures for the present study included: (1) the duration of the /h/, (2) the presence or absence of voicing during the /h/, and (3) the formant structure (F_1-F_2-F_3) for the /h/ measured at the midpoint of the segment, and (4) F_1-F_2-F_3 for the following vowel nucleus, also measured at its midpoint. Voicing during an /h/ segment was defined as the presence of a periodic component in the glottal signal. Ten of the 11 word pairs on the test were analyzed, including: *hair-air, hate-ate, and-hand, old-hold, eat-heat, arm-harm, add-had, ail-hail, all-hall,* and *hat-at.* The one excluded word pair was misread by several of the speakers during the original recording *(ash-hash).* One hundred and fourteen of the 385 tokens (11 tokens × 35 speakers) for this contrast were classified as errors based on listener responses in the intelligibility test.

Reliability

Two words for each contrast and speaker were remeasured by the investigator and also by a second investigator to obtain estimates of intra- and interjudge reliability. The overall correlation coefficient for intrajudge reliability was 0.97 across measures and speakers, for estimates of interjudge reliability the value was 0.93. Two sample t tests showed no significant differences between the three sets of acoustic measurements (original vs. intrajudge and original vs. interjudge). Standard errors of measurement were computed for each speaker group; values, expressed as percentages, were less than 3% in all cases, suggesting that none of the groups was more challenging in terms of acoustic measurement. Reliability for the two measures that required the investigator to make a categorical judgment (presence or absence of voicing in the initial /h/ segment and voicing during the stop closure) was 0.86 and 0.79 for intra- and interjudge measures, respectively.

Results

Syllable-Initial Voiced-Voiceless Contrast

Tokens for the initial voicing contrast included 6 targets with a voiced initial consonant and 3 with a voiceless initial consonant. Thirty-seven of the 47 voicing errors were perceived as voiceless when the intended target was voiced and 10 of the errors were perceived as voiced when the intended target was voiceless.

Voicing

For the tokens with a voiceless word-initial consonant, the presence of voicing during the closure interval was noted in 17 tokens produced by the male speakers and 3 produced by the female speakers. The listeners perceived all 20 of these tokens as targets (i.e., voiceless). The duration of the voicing ranged from 51 to 120 ms. Only 1 NG-M showed evidence of voicing for a target token perceived as voiceless; speakers in the disorder group produced the remaining 19 tokens. For all of the error tokens (perceived as voiced when the target was voiceless) evidence of voicing prior to the burst was seen. For tokens where the target contained a voiced initial consonant and were perceived as such, 2 tokens produced by the female speakers had no evidence of voicing. The 37 error tokens (perceived as voiceless) all contained evidence of voicing. The mean duration of voicing during the closure interval was essentially identical for error and target tokens (mean = 94.8 ms and mean = 93.7 ms, respectively).

Voice Onset Time

The VOT distributions for target and error tokens can be found in figure 1, where data are presented in two panels, for male and female speakers, respectively. Data were divided by gender to ensure that there were no gender effects, as Kent et al. [11] reported voicing-related gender effects. The target tokens in the plots are represented by filled shapes and the errors by unfilled shapes. Within each panel the top three rows show cases where the target was voiced and the error voiceless; the bottom five rows show the opposite cases. Thus for the top three rows,

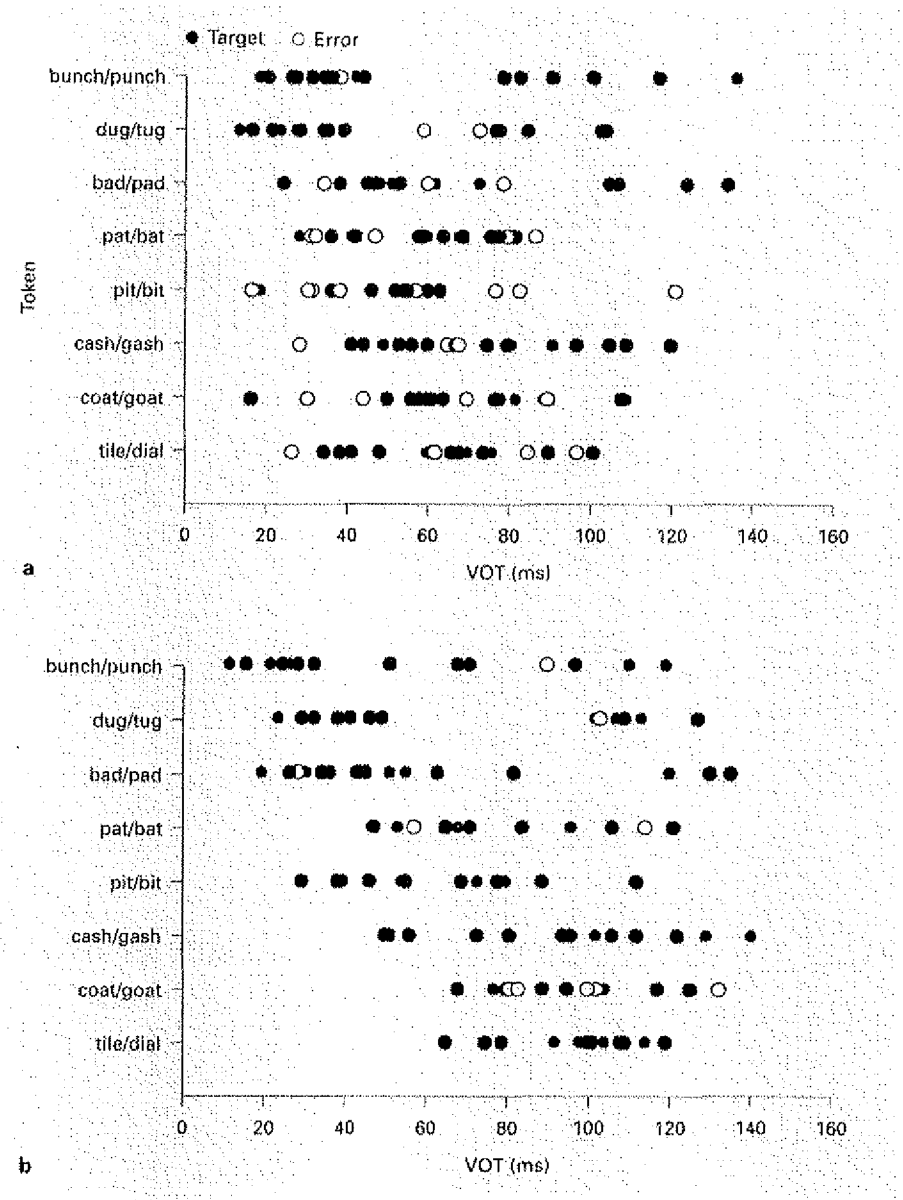

Fig. 1. The range of VOT measured for the male (**a**) and female (**b**) speakers for each word pair. Filled shapes represent target tokens and unfilled errors.

errors (perceived as voiceless tokens) might be expected to occur toward the right of the distribution (longer VOTs) compared to the tokens perceived as targets (voiced). For the bottom five rows of the graph, errors (voiced) might be expected to occur to the left of the distribution (shorter VOTs) compared to the corresponding voiceless target tokens.

Measures of VOT for the normal geriatric speakers (across gender) on target tokens were

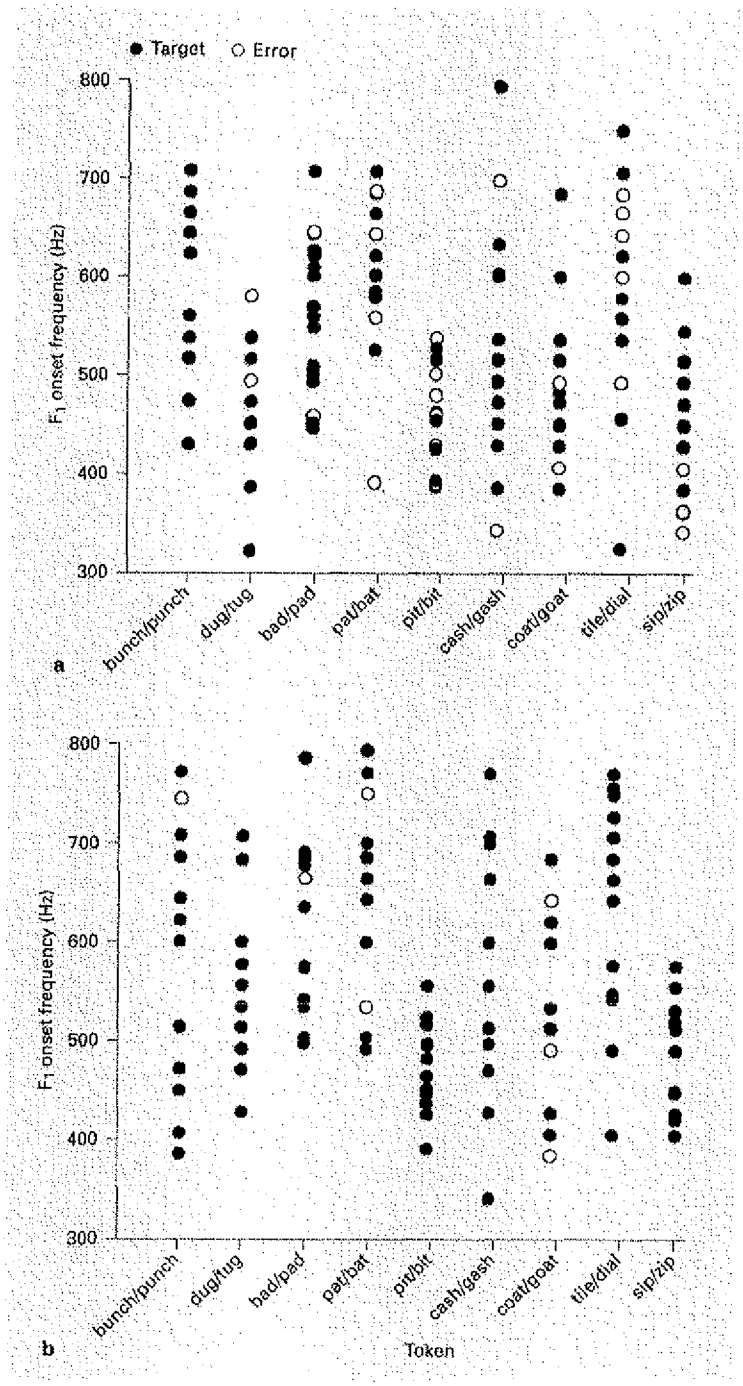

Fig. 2. F_1 onset values for male (**a**) and female (**b**) speakers. The target tokens with voiced initial consonants are shown to the left in the plot. The filled shapes represent target tokens and the unfilled shapes error tokens.

not always consistent with expectations. For example, for several speakers VOTs measured for the voiced targets were in excess of 60 ms (solids in the top three rows of fig. 1) and corresponding error tokens (perceived as voiceless) did not move in the expected direction. Comparisons between the NG and disorder groups show that for both genders, the VOTs for the disorder groups overlap entirely with variability of the target productions for the NG speakers. For the male speakers in the disorder groups, however, a larger number of the error tokens (perceived as voiced) fell in the lower half of the VOT range measured for the voiceless target tokens. Although there are VOT measurements consistent with expectations, some tokens did not conform to predicted VOTs based on their perception or data in the literature (i.e., tokens perceived as voiced did not always have shorter VOTs than their voiceless counterparts).

F_1 Onset

F_1 onset values measured at the first glottal pulse of the vowel are shown in figure 2 for the male and female speakers. The first three word pairs starting at the left in the plots contain voiced targets while all others contain voiceless initial stop targets. The filled shapes represent the target tokens and the unfilled represent the error tokens. It was expected that the F_1 onset values for vowels which follow a voiced consonant would be lower compared to their voiceless cognate [35]; however, F_1 onset values measured for both the target and error tokens in the present study showed extensive overlap across gender and speaker group.

Vowel Duration

Voicing does have a recognized effect on vowel duration, i.e., a vowel, regardless of identity, is longer when it is juxtaposed with a voiced stop compared to a voiceless stop [31], although there is less of an effect when the stop is prevocalic than postvocalic. Comparisons of the vowel durations for the error tokens to those measured in target productions showed no significant differences (Mann-Whitney test, W = 7,621, p = 0.7356). The majority of speakers had shorter vowel durations in the erred tokens compared to the corresponding target tokens (38 of 47 error tokens), with differences ranging between 10 and 60 ms. Six of the 9 error words had voiced initial consonants; therefore, the finding of shorter vowel duration was inconsistent with predictions regarding vowel length. In the few cases where error tokens had longer vowel durations compared to target tokens (9 of 47 error tokens), as was expected for voiced or voiceless errors, the differences were less than 15 ms.

Syllable-Initial /h/-Vowel Contrast

Expectations for the acoustic characteristics of tokens perceived as having an initial /h/ were a separate and identifiable noise segment characterized by high-energy noise, and possibly a formant structure matching that of the adjacent vowel with an absence of, or weak F_1. Another possible characteristic of the /h/ segment would be the presence of two separate formants in the area of F_2 which merge into one for the following vowel segment. For tokens perceived as having an initial vowel only (target or error), a vowel segment with a formant structure consistent with the vowel identity should be identifiable on the spectrogram. In these tokens, none of the characteristics of an /h/ listed above should be apparent on the spectrogram. Three of the target words had an initial vowel only *(air, ate,* and *at),* so for these tokens perception of *hair, hate,* and *hat* represented one type of /h/ -vowel error. The remaining 7 tokens had an

Table 3. Number of subjects within each gender who voiced the initial /h/ segment as evidenced by the presence of a voice bar and the corresponding perception of the token

Token	Number of tokens in each category		Number voiced	
	male	female	male	female
Target /h/ initial	65	95	22	3
Error vowel initial	75	10	15	4
Target vowel initial	35	41	0	0
Error /h/ initial	25	4	19	8

Total possible tokens in each category were also listed.

initial /h/ as the target (e.g., *hail*) and perception of the initial vowel only (e.g., *ail*) was considered an error. Errors, in both directions, were more frequent for male speakers compared to females (88 versus 12%).

Initial /h/ Voicing

Voicing for the initial /h/ was defined as the presence of a voice bar during an initial noise segment. The number of subjects within each gender group for whom a voice bar was present during the initial /h/ segment is listed in table 3 as well as the total number of tokens analyzed in each category. There were additional tokens where an initial /h/ segment was identifiable (by the formant structure or presence of noise), but the segment contained no evidence of a voice bar; these tokens were not counted in the table. For the male speakers, target tokens containing an initial /h/ were more frequently voiced than in females. A number of errors whose targets contained initial /h/ but were perceived with an initial vowel only (e.g., *add, all, arm, old*) contained a voice bar during a separate and identifiable /h/ noise segment. This pattern was also seen primarily for male speakers. Evidence of voicing was also found for the 3 tokens whose intended target was vowel-only yet were perceived with an initial /h/ (e.g., *hair, hat*).

Initial /h/ Duration

The duration of the initial /h/ segment was measured in both target and error tokens when they contained clear acoustic evidence of an /h/-like noise segment. It was expected that in targets and errors where an initial vowel was perceived, no /h/-like segment would be seen. In several cases, as mentioned above, evidence of an initial /h/ was observed in these tokens. The mean durations and standard deviations of the /h/ segment are shown in tables 4 and 5 for male and female speakers, respectively. Data are shown in separate columns for each disorder group by word with the upper panel of the table showing data for target tokens and the lower panel showing data for the corresponding error tokens. Blanks within the table indicate that the measure was not appropriate (e.g., no /h/ segment was measured for targets containing an initial vowel in the upper panel) or no error tokens were found for a given word pair (lower panel). The length of the /h/ noise segment appeared more variable across speakers and tokens in the ALS and CVA disorder groups compared to NG speakers. Typical durations for the /h/ segment in the NG males were below 100 ms for the target tokens (table 4, upper panel, 6 of 7 cases). These values were slightly higher for the NG females, but were

Table 4. Duration means and SDs for the initial /h/ segment in target and error tokens produced by male speakers

	NG-M		ALS-M		PD-M		CVA-M	
	mean	SD	mean	SD	mean	SD	mean	SD
Target								
air								
at								
ate								
had	83.4	24.5	279.5	41.7	68.5	16.3	60	4.24
hail	77	42.4	116.3	79.6	80.5	24.7	59.7	20.6
hall	107.2	38.5	235	158	110		87.7	42.4
hand	75.4	14.71	132.3	113.8	91	1.41	72.33	4.93
harm	84.3	24	133.7	105.1	114	17	55	
heat	69	8.12	94.3	56	150		54	
hold	80	5.77	53.5	10.61	84		95	
Error								
hair	102		130	72.6	143.4	44.4		
hat	213		103.3	114.2	140	35.7	252.33	9.29
hate			56		135	81.9	80.5	64.3
add			60.7	29.7	78.7	35.4	59.7	40.2
ail	38		39	21.2	74	27.1	61	15.6
all			72.3	53.8	95.5	46.5	130.5	27.6
and			35	28.3	117	60.8	92	2.83
arm	83.5	9.19	41	185	87.3	67.6	82.5	22.4
eat	78		59	83.4	54.5	35.9	80.8	21.8
old	321		123.3	60.4	58	13.64	163	115.3

Blanks within the table indicate no measurements were made.

shorter than 133 ms in all cases. Mean target values for the /h/ duration in the male disorder groups (except CVA-M) were longer than NG for the target tokens (group means NG-M = 82.2 ms; ALS-M = 149.2 ms; PD-M = 99.8 ms; CVA-M = 69 ms). For the females, the opposite was true (group means NG-F = 114 ms; ALS-F = 98 ms; PD-F = 94 ms).

A comparison of error segment durations for the NG speakers and disorder groups showed that error tokens consistently had longer segment durations compared to the corresponding target tokens for the female disorder groups (mean = 102.7 ms for targets, mean = 182.5 ms for errors). Two exceptions were the *harm-arm* pair for the PD-F group and *heat-eat* for the ALS-F group; in both these cases the error token had shorter segment durations compared to the target tokens for speaker within the group. For the ALS-M groups, however, durations were always shorter with only a single exception *(hold-old)*, (table 4). Durations for the other two male disorder groups ranged from 54 to 250 ms and were similar to values measured for the corresponding target tokens.

Table 5. Duration means and SDs for the initial /h/ segment in target and error tokens produced by female speakers

	NG-F		ALS-F		PD-F	
	mean	SD	mean	SD	mean	SD
Target						
air						
at						
ate						
had	132.4	44.6	87.4	19.98	66	11.31
hail	111.8	20.9	87.4	23	77.4	20.85
hall	86.4	19.42	86.7	25.6	192	258
hand	112.6	13.9	88.4	6.88	66	15.77
harm	114.8	35.5	86.2	49.2	92.67	8.14
heat	117.8	23.5	98.7	37.3	83	5.66
hold	123.4	23	154.2	135.4	78	24.3
Error						
hair			162.5	46	174.7	61.1
hat			257		244	31.1
hate					137	
add					124	71.6
ail						
all			304			
and					130	
arm					77.5	7.78
eat			88		95.7	69.8
old					158.5	6.36

Blanks within the table indicate no measurements were made.

/h/ Formant Structure

The formant structure similar to the neighboring vowel was clearly visible in most cases where an initial /h/ was expected (targets) or where one was perceived (errors). No evidence of a double formant in the area of F_2 was present in any of the tokens measured. F_1, which was expected to be weak or absent in an /h/ segment, was present in about half the target tokens with an initial /h/ produced by the NG speakers (5/7 for males and 4/7 for females). Values for F_1 were higher, typically between 100 and 200 Hz, than those for the adjacent vowel measured at the midpoint of the vowel nucleus. In a few cases, the second and third formants were also slightly higher than those measured for the adjacent vowel (e.g. *had, heat*). For the disorder groups, the first formant was present for many of the target tokens which contained an initial /h/ (78 of 175 tokens). The formant frequency values (F_1) were more typical of those expected for adjacent vowel than those measured for the NG speakers, and were very similar between groups. The same was true for F_2 and F_3, although values were slightly lower than those measured for the vowel (a difference of less than 150 Hz). Almost all target tokens perceived with an initial /h/ aperiodic contained high-energy noise above 3,500 Hz (NG and disorder groups).

Comparisons of the data for target and error tokens in cases where the target contained an initial /h/ but listeners perceived only a vowel showed that an /h/ segment was often identifiable and similar in structure to those measured for target tokens perceived with an /h/.

Discussion

The examination of a contrast addressing the potential effects of voicing errors on intelligibility is motivated primarily by perceptual data as well as clinical intuition, rather than any clear evidence that these kinds of control issues are typical in the motor speech disorder population exclusively. In fact, there is evidence that partial loss of voicing control may be associated with aging in general rather than a disease process [42]. Increases in stiffness of muscle tissue, due to thinning and increased connective tissue, may result in decreased efficiency of muscle contraction and thus, reduced abductory force for the laryngeal devoicing gesture, especially in elderly males. In the present study, male speakers (across groups) tended to have relatively higher error rates for the glottal-null contrast and to a lesser degree, for the initial voicing contrast. Women, on the other hand, had relatively little difficulty with either of these phonetic contrasts.

Acoustic analysis was used in the present study to examine the relationship between 'cues' and perception of the phonetic target for an initial voicing and glottal-null contrast. The data from the present study do not support a simple relationship between the acoustic measures of segmental laryngeal control, presumed to be the signal basis of the voicing distinction for English stops and fricatives, and the segmental voicing errors derived from a perceptual test.

Performance of the NG Speakers

It is important to look at performance of the control group to determine if the acoustic correlates selected for each contrast were comparable to those reported in the literature, especially since much of acoustic-phonetics research has focused on the speech of young, healthy individuals. The few studies that have included older individuals have shown various anatomical and physiological changes resulting from aging, which have some general effects on speech production. In general, acoustic measures for the target tokens were largely consistent with data available in the literature [26], suggesting that comparisons between performance of the disorder groups and previously reported data are appropriate. One exception was the VOT measure. Increased variability in the VOT measure, which is reflective of a coordinated activity that is apparently a sensitive indicator of aging in the speech production system, was comparable to previously reported data for aged males [22, 23, 42]. However, the absolute VOT values measured for the voiced targets were extremely long (~ 90–100 ms), in some cases double the expectation. Reasons for this discrepancy are not clear. Values for the VOTs of voiced stops available in the literature are typically from stops in carrier phrases, whereas the current data are from single words of the Kent et al. [14] intelligibility test. In English, it is not unusual for utterance-initial voiced stops to be produced without voicing during the closure interval, since there is no voiced-voiceless unaspirated contrast for stops; however, the very long values of VOTs for some voiced stops are clearly different from expectations. These unusual values, however, do not change the overall picture of pervasive overlap of VOT values in the error analysis, which suggests that VOT measures do not 'explain' the perceptual errors.

Performance of the Disordered Speakers

For the voiced-voiceless contrast, evidence of voicing during the closure interval for intended voiceless stops would be thought to increase the likelihood that the token would be identified erroneously. Long voicing durations during the closure interval for voiced stops should, on the other hand, not affect the percept of the token. The present study, however, found that in 1 case for the NG speakers and 19 cases for the disorder group, voicing was present during a closure interval of an intended voiceless stop, yet listeners still perceived the token as intended. Likewise, in 2 tokens intended and perceived as voiced targets no evidence of voicing was found. This suggested that the cue of voicing does not appear to be directly related to the listeners' response for each token. In these cases, it may be that ambiguous or conflicting cues in the signal forced listeners to alter their perceptual strategies. Lisker [43] has suggested that multiple acoustic events cue stop voicing simultaneously and it is the integration of these cues that allows a listener to arrive at a given percept. Based on this discussion, the apparent mismatch between a single cue and the listeners' percept may not be unexpected.

There are data in the literature indicating that older speakers have significantly shorter VOTs than younger speakers [23, 24, 44 45]. Weismer [21] reported that VOTs in PD were substantially shorter than they were in elderly, neurologically healthy men (aged 65–82 years). This was interpreted as reflecting the exaggeration of aging effects of the larynx that may accompany PD. Most of these studies have been done on male speakers, but at least two comparisons of older and younger female speakers [23, 46] did not find any age-related VOT differences, raising the possibility that the effect is gender-specific. In the present data, VOT values measured in words with an initial voiced stop for the normal geriatric groups (both male and female) were longer than those reported in the literature [26].

Difficulty with voicing control for the glottal-null contrast was seen as the presence of a voice bar during a perceived voiceless segment. This was found for 25 correct tokens and 27 error tokens all perceived as having an initial /h/ (table 3). Based on these findings, it seems that the voicing cue did not correlate with the listeners' percept of the tokens. There are several possible explanations for the mismatch. First, voicing during an expected voiceless segment may be indicative of a normal change related to the aging process. Listeners may develop a tolerance for such information in older speakers and do not utilize it to make a decision. Or, other cues were strong enough to override the voicing. Similar production characteristics for a younger speaker may or may not be interpreted the same way. Weismer [21] also discussed varying perceptual expectations of speech dimensions, depending on a speaker's inferred age. Since Lehiste's [40] study of a word-initial /h/ was based on only 1 speaker, the phonetic expectation may also not have been accurate. Yet another explanation may be that because intervocalic /h/s are often partially or completely voiced, listeners may be more forgiving of an apparently voiced /h/ in a word-initial position. Because of similarities in formant structure for an /h/ and neighboring vowel, listeners may use the noise spectrum to reach a decision on the presence or absence of an /h/. Any of these cases may be true, since representation of the contrast by this acoustic variable did not capture the essential acoustic cues for the contrast in the present study.

The variability across speakers within a single disorder group was larger than has been reported for younger speakers. Cariski [42] has also reported an increased variability in VOT measured for older, healthy speakers compared to young speakers, but did not find any significant differences for the range when comparing young and older speakers. VOTs measured in the present study for the error tokens largely overlapped with those measured for the target tokens, regardless of the direction of the error (fig. 1). Differences in VOT between the disorder groups were not found in the present data [26]. In addition, there were no clear differences between VOT for the disorder groups compared to the normal geriatric speakers. Evidence of conflicting cues in the acoustic signal was also found for the glottal-null contrast across speaker groups. The presence of a voice bar throughout the duration of the word-initial /h/ segment, as found in the present study, has never been described in the literature. The speakers who produced the segment with voicing tended to be male (97 of 143 cases). We can predict that the presence of a voice bar makes it more likely that listeners would mark the token as having an initial vowel. However, spectrographic evidence of high-energy noise and weak F_1 in the initial segment was also found in these tokens. These latter cues should have provided cues for the listeners which identified the token as having an initial /h/. Error proportions for tokens with both a voice bar and turbulent noise ranged from 0.4 to 0.7 (4 of 10 listeners and 7 of 10 listeners, respectively); so, as a group it appears that the listeners did not agree on the identity of the token, further suggesting that they may have had difficulty with conflicting cues [26, 28].

One concern about the glottal-null contrast, which was not considered in the listening portion of the experiment, was differences in consonant discrimination as a function of presentation level. Kent et al. [47] showed that performance intensity functions for individual consonants varied widely. Even at the highest presentation level, 60 dB, several weak consonants were not identified with 100% accuracy. This included the word-initial /h/. Therefore, it is possible that the listeners in the present study did not receive all cues found in the acoustic signal for the /h/. In particular, the high-frequency noise that was weak compared to the formants may have been missed.

Conclusions

Laryngeal behavior for segmental distinctions is often disrupted in motor speech disorders [15, 48]. Studies of speech intelligibility have demonstrated that obstruent voicing contributes significantly to speech intelligibility deficits [11, 29]. While these listener-based profiles provide some strong evidence which points to laryngeal dysfunction as a primary clinical issue, several studies have reported significant changes in laryngeal structure and function associated with aging in general, especially for males. Weismer [21] and Liss et al. [24] have suggested that studies of changes in articulatory function associated with aging may be important in understanding neurologic disease; e.g., investigations of VOT in PD suggest that these speakers may represent exaggerated effects of aging on laryngeal function. The relatively high error rates for laryngeal contrasts in the present study led to the examination of the acoustic records to identify the correlates of the phonetic impairment. Results suggest that the problem appears to be one with timing of laryngeal adjustment. Voicing was often initiated early, creating confusion about the intended voicing feature. For example, voicing began during what should have been a voiceless initial segment. Similarities in the acoustic characteristics of the target and error tokens for the NG and disorder groups on the initial /h/ segment also suggest that there may in fact be specific problems in laryngeal control that are due to aging and that these characteristics may be exaggerated in neurological populations.

A primary goal in developing a model of speech intelligibility in dysarthria is to examine how specific changes in production reflected in the acoustic signal affect the perception of a given token and ultimately overall speaker intelligibility. Results of this study along with those reported by Bunton and Weismer [28] suggest that research protocols need to utilize a more multidimensional approach, whereby a priori acoustic cues are not targeted but differences in the acoustic signal across a contrast are examined. This type of an approach will allow researchers to explore the many cues contained in the acoustic signal and is similar to strategies used by listeners who most likely integrate a set of different cues to arrive at a given percept, with primacy assigned to the integrated percept and not any of the contributing components. Examination of the multiple cues in acoustic signals and exploration of different weights that may be assigned to these cues may offer insight into the articulatory-acoustic-auditory relationship.

Acknowledgments

This work was supported in part by NIH R01 DC00319 and T32 DC00042.

References

1 Kent RD: Research needs in the assessment of speech motor disorders; in Cooper J (ed): Assessment of Speech and Voice Production: Research and Clinical Applications. NIDCD Monogr Bethesda, National Institute of Deafness and Other Communication Disorders, 1992, pp 17–29.
2 Ramig L: The role of phonation in speech intelligibility: A review and preliminary data from patients with Parkinson's disease; in Kent R (ed): Intelligibility in Speech Disorders: Theory, Measurement, and Management. Amsterdam, Benjamins, 1992, pp 119–155.
3 McGarr NS, Osberger MJ: Pitch deviancy and intelligibility. J Commun Disord 1978;11:237–248.
4 Yorkston KM, Buckelman DR: Assessment of Intelligibility of Dysarthric Speech. Tigard, C.C. Publications, 1981.
5 Tikofsky RS: A revised list for the estimation of dysarthric single word intelligibility. J Speech Hear Res 1970;13:59–64.
6 Darley FL, Aronson AE, Brown JR: Differential diagnostic patterns of dysarthria. J Speech Hear Disord 1969;12:246–269.
7 Darley FL, Aronson AE, Brown JR: Cluster of deviant speech dimensions in the dysarthrias. J Speech Hear Disord 1969;12:462–496.
8 Smith CR: Residual hearing and speech production in deaf children. J Speech Hear Res 1975;18:795–811.
9 Weismer G, Kent RD, Hodge M, Martin R: The acoustic signature for intelligibility test words. J Acoust Soc Am 1988;84:1281–1291.
10 Samar VJ, Metz DE: Criterion validity of speech intelligibility rating-scale procedures for the hearing-impaired population. J Speech Hear Res 1988;31:307–316.
11 Kent R, Kim HH, Weismer G, Kent JF, Rosenbek J, Brooks BR, Workinger M: Laryngeal dysfunction in neurological disease: Amyotrophic lateral sclerosis, Parkinson disease, and stroke. J Med Speech Lang Pathol 1994;2:157–175.
12 Yorkston KM, Buckelman DR: Clinical Management of Dysarthric Speakers. Boston, Little, Brown, 1988.
13 Kent RD, Weismer G, Kim HH: Segmental and suprasegmental laryngeal dysfunction in neurological disease: Amyotrophic lateral sclerosis, Parkinson's disease, stroke, and cerebellar disease. Conf Neurol Disord Laryng Function, Los Angeles, 1980.
14 Kent RD, Weismer G, Kent JF, Rosenbek JC: Toward explanatory intelligibility testing in dysarthria. J Speech Hear Disord 1989;54:482–499.
15 Weismer G: Control of the voicing distinction for intervocalic stops and fricatives: Some data and theoretical considerations. J Phonet 1980;8:427–438.
16 Farmer A: Voice onset time in cerebral palsied speakers. Folia Phoniatr Logop 1980;32:267–273.

17 Kent JF, Kent RD, Rosenbek J, Weismer G, Martin R, Sufit R, Brooks B: Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis. J Speech Hear Res 1992;35:723–733.

18 Logemann JA, Fisher HB: Vocal tract control in Parkinson's disease: Phonetic feature analysis of misarticulation. J Speech Hear Disord 1978;46:348–352.

19 Platt LJ, Andrews G, Young M, Quinn PT: Dysarthria of adult cerebral palsy. 1. Intelligibility and articulatory impairment. J Speech Hear Res 1980;23:28–40.

20 Weismer G, Martin R: Acoustic and perceptual approaches to the study of intelligibility; in Kent R (ed): Intelligibility in Speech Disorders: Theory, Measurement, and Management. Amsterdam, Benjamins, 1992, pp 68–118.

21 Weismer G: Articulatory characteristics of Parkinsonian dysarthria; in McNeil MR, Rosenbek JC, Aronson AE (eds): The Dysarthrias: Physiology, Acoustics, Perception, Management. San Diego, College-Hill Press, 1984, pp 101–130.

22 Weismer G, Liss JM: Age and speech motor control; in Ripich D (ed): Handbook of Aging and Communication. San Diego, College-Hill Press, 1992, pp 205–225.

23 Weismer G, Fromm D: Acoustic analysis of geriatric utterances: Segmental and nonsegmental characteristics which relate to laryngeal function; in Bless DM, Abbs JH (eds): Vocal Fold Physiology: Contemporary Research and Clinical Issues. San Diego, College-Hill Press, 1983, pp 317–332.

24 Liss JM, Wiesmer G, Rosenbek JC: Selected acoustic characteristics of speech production in very old males. J Gerontol B Psychol Sci Soc Sci 1990;45:35–45.

25 Ziegler W, Hartmann E, von Cramon D: Word identification testing in the diagnostic evaluation of dysarthric speech. Clin Linguist Phonet 1988;2:291–308.

26 Bunton K: An Acoustic Item-Analysis of Speech Intelligibility, unpublished doct diss, University of Wisconsin-Madison, 1999.

27 McQueen J, Cutler A: Cognitive processes in spoken-word recognition; in Hardcastle WJ, Laver JDMH (eds): The Handbook of Phonetic Sciences. Oxford, Blackwell, 1997, pp 566–585.

28 Bunton K, Weismer G: The relationship between perception and acoustics for a high-low vowel contrast produced by dysarthric speakers. J Speech Lang Hear Res 2001;44:1215–1228.

29 Kent RD, Kent JF, Weismer G, Sufit R, Rosenbek JC, Martin RE, Brooks BR: Impairment of speech intelligibility in men with amyotrophic lateral sclerosis. J Speech Hear Disord 1990;55:721–728.

30 Kewley-Port D, Pisoni DB, Studdert-Kennedy M: Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. J Acoust Soc Am 1983;73:1779–1793.

31 Klatt DH: Voice onset time, frication and aspiration in word-initial consonant clusters. J Speech Hear Res 1975;18:686–706.

32 Lisker L, Abramson AS: A cross language study of voicing of initial stops: Acoustic measurements. Word 1964;20:384–422.

33 Pickett JM: The Sounds of Speech Communication. Baltimore, University Park Press, 1980.

34 Stevens KN, Blumstein SE: Invariant cues for place of articulation in stop consonants. J Acoust Soc Am 1978;64:1358–1368.

35 Kluender KR: Speech perception as a tractable problem in cognitive science; in Gernsbacker MA (ed): Handbook of Psycholinguistics. San Diego, Academic Press, 1994, pp 173–217.

36 Miller GA, Nicely PE: An analysis of perceptual confusions among some English consonants. J Acoust Soc Am 1955;27:338–352.

37 Massaro D, Cohen M: The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. J Acoust Soc Am 1976;60:704–717.

38 Milenkovic P: Cspeech. University of Wisconsin-Madison, Department of Electrical and Computer Engineering, 1994.

39 Olive JP, Greenwood A, Coleman JS: Acoustics of American English Speech: A Dynamic Approach. New York, Springer, 1993.

40 Lehiste I: Acoustic Characteristics of Selected English Consonants. Bloomington, Indiana University, 1964.

41 Fant G: Acoustic Theory of Speech Production. The Hague, Mouton, 1960.

42 Cariski D: Effects of Speaking Rate and Contrastive Stress on Acoustic Measurements of Prosody: unpublished doct diss, University of Wisconsin-Madison, 1992.

43 Lisker L: The pursuit of invariance in speech signals. J Acoust Soc Am 1985;77:1199–1202.

44 Smith B, Wasowicz J, Preston J: Temporal characteristics of the speech of normal elderly adults. J Speech Hear Res 1987;30:522–529.

45 Sweeney P, Baken R: Voice onset time in a normal aged population. J Speech Hear Res 1982;25:129–134.

46 Neiman G, Kluch R, Shuey E: Voice onset time in young and 70-year-old women. J Speech Hear Res 1983;26:118–123.

47 Kent RD, Wiley TL, Strennen ML: Consonant discrimination as a function of presentation level. Audiology 1979;18:212–224.

48 Lofqvist A: Interarticulator programming in stop production. J Phonet 1980;8:475–490.