

# Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners<sup>a)</sup>

Sarah Hargus Ferguson<sup>b)</sup>

Department of Communication Sciences and Disorders, University of Utah, 390 South 1530 East, Room 1201, Salt Lake City, Utah 84112

Hugo Quené

Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands

(Received 15 May 2013; revised 7 March 2014; accepted 9 April 2014)

The present investigation carried out acoustic analyses of vowels in clear and conversational speech produced by 41 talkers. Mixed-effects models were then deployed to examine relationships among acoustic and perceptual data for these vowels. Acoustic data include vowel duration, steady-state formant frequencies, and two measures of dynamic formant movement. Perceptual data consist of vowel intelligibility in noise for young normal-hearing and elderly hearing-impaired listeners, as reported by Ferguson in 2004 and 2012 [J. Acoust. Soc. Am. **116**, 2365–2373 (2004); J. Speech Lang. Hear. Res. **55**, 779–790 (2012)], respectively. Significant clear speech effects were observed for all acoustic metrics, although not all measures changed for all vowels and considerable talker variability was observed. Mixed-effects analyses revealed that the contribution of duration and steady-state formant information to vowel intelligibility differed for the two listener groups. This outcome is consistent with earlier research suggesting that hearing loss, and possibly aging, alters the way acoustic cues are used for identifying vowels. © 2014 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4874596>]

PACS number(s): 43.70.Fq, 43.71.Es, 43.71.Ky [PBN]

Pages: 3570–3584

## I. INTRODUCTION

Aural rehabilitation programs for individuals with hearing loss typically include training in strategies designed to improve communication. One common communication strategy is to request that one's communication partner speak more clearly. Several laboratory studies have demonstrated that complying with this request should make talkers easier to understand. In most of these studies, talkers read written speech materials aloud in two conditions. In the first, they are asked to speak as they do in ordinary conversation while in the second they are instructed to speak as though talking to someone who has difficulty understanding them. These two conditions have typically been called “conversational” and “clear” speech, respectively.<sup>1</sup> When the materials are subsequently presented to listeners for identification, a “clear speech benefit” has been found for several listener groups, including adults with sloping high-frequency sensorineural hearing loss (e.g., Schum, 1996; Ferguson, 2012), adult cochlear implant users (Liu *et al.*, 2004; Ferguson and Lee, 2006), children with learning disabilities (Bradlow *et al.*, 2003), and normal-hearing listeners identifying the stimuli in noise and/or reverberation (e.g., Payton *et al.*, 1994; Ferguson, 2004; Smiljanic and Bradlow, 2005). However, asking a talker to speak clearly does not always produce a

clear speech benefit. In studies with multiple talkers, the magnitude of the clear speech effect has varied widely among talkers (Gagné *et al.*, 1994; Schum, 1996; Ferguson, 2004; Ferguson, 2012).

Acoustic analyses have revealed numerous differences between clear and conversational speech. Talkers speaking clearly speak more slowly (Picheny *et al.*, 1986; Bradlow *et al.*, 2003; Smiljanic and Bradlow, 2005; Smiljanic and Bradlow, 2008; Ferguson *et al.*, 2010; Rosen *et al.*, 2011; Lam *et al.*, 2012) and more loudly (Picheny *et al.*, 1986; Ferguson *et al.*, 2010; Hazan and Baker, 2011;<sup>2</sup> Lam *et al.*, 2012) and with a higher voice pitch (Bradlow *et al.*, 2003; Hazan and Baker, 2011). Talkers also use a more variable voice pitch in clear speech than in conversational speech (Picheny *et al.*, 1986; Bradlow *et al.*, 2003; Hazan and Baker, 2011). These global clear speech acoustic changes are accompanied by phoneme-level changes such as stronger final consonants (Picheny *et al.*, 1986; Bradlow *et al.*, 2003; Ferguson *et al.*, 2010) and vowel modifications including an expanded vowel space (Picheny *et al.*, 1986; Ferguson and Kewley-Port, 2002; Bradlow *et al.*, 2003; Smiljanic and Bradlow, 2005; Ferguson and Kewley-Port, 2007; Ferguson *et al.*, 2010; Hazan and Baker, 2011; Lam *et al.*, 2012), greater dynamic formant movement (Ferguson and Kewley-Port, 2002; Ferguson and Kewley-Port, 2007; Lam *et al.*, 2012), and longer vowel durations (Ferguson and Kewley-Port, 2002; Ferguson and Kewley-Port, 2007; Picheny *et al.*, 1986; Lam *et al.*, 2012). Similar vowel modifications have been observed in studies comparing “hyperarticulated” and citation-style speech (e.g., Moon and Lindblom, 1994).

<sup>a)</sup>Portions of this work were presented at the 164th meeting of the Acoustical Society of America, Kansas City, Missouri, 2012.

<sup>b)</sup>Author to whom correspondence should be addressed. Electronic mail: sarah.ferguson@hsc.utah.edu

Although both the intelligibility and the acoustic differences between clear and conversational speech are well-established, the relationship between them is not. The present study aims to elucidate this relationship somewhat by investigating the acoustic characteristics that underlie vowel intelligibility in clear and conversational speech for young normal-hearing (YNH) and elderly hearing-impaired (EHI) listeners. It continues a line of research begun by [Ferguson and Kewley-Port \(2002\)](#) which has two distinguishing characteristics. First, it focuses on a single phoneme category in a fixed context, thus limiting the enormous pool of potentially important clear speech acoustic changes to a more manageable set. Second, it takes advantage of naturally-occurring variability in speech production. In [Ferguson and Kewley-Port \(2002\)](#), the variance exploited was among multiple productions by a single talker (intra-talker variability), while [Ferguson and Kewley-Port \(2007\)](#) used the differences found among 10 talkers selected from the 41-talker Ferguson Clear Speech Database (inter-talker variability; [Ferguson, 2004](#)). Using vowel intelligibility data from YNH listeners, [Ferguson and Kewley-Port \(2007\)](#) compared the clear speech acoustic changes observed in a group of five talkers who had shown a large clear speech benefit in vowel intelligibility to those observed in a group of five talkers who had produced no clear speech benefit. These “extreme groups” comparisons suggested that greater increases in vowel duration and greater expansion of the vowel formant space in clear speech were associated with greater increases in vowel intelligibility, while increases in dynamic formant movement were not.

The current project extended the work of [Ferguson and Kewley-Port \(2007\)](#). Vowel acoustic data have now been obtained from all 41 talkers in the Ferguson Database and are reported here, as is the relationship between these acoustic data and vowel intelligibility in noise data for both EHI listeners (reported in [Ferguson, 2012](#)) and YNH listeners (reported in [Ferguson, 2004](#)). Both perceptual studies revealed considerable inter-talker variability both for vowel intelligibility in each speaking style and for the magnitude of the clear speech vowel intelligibility effect. In addition, [Ferguson \(2012\)](#) found that although better signal-to-noise ratios (SNRs) in the intelligibility test procedure yielded higher overall vowel intelligibility for EHI listeners (tested at  $-3$  dB SNR) than for the YNH listeners (tested at  $-10$  dB SNR), the two groups benefited equally from clear speech. Furthermore, it was generally the case that talkers who produced a clear speech benefit in vowel intelligibility for YNH listeners also produced a benefit for EHI listeners. This agrees with previous studies showing similar clear speech effects for listeners with hearing loss and listeners with normal hearing (e.g., [Payton et al., 1994](#)), but diverges from [Ferguson and Kewley-Port \(2002\)](#), whose single talker produced a significant clear speech vowel intelligibility benefit for YNH listeners, but no benefit for EHI listeners.

Regression analyses performed within individual vowel categories in [Ferguson and Kewley-Port \(2002\)](#) indicated that although both YNH and EHI listeners used all three of the traditional vowel acoustic cues (steady-state formant frequencies, dynamic formant movement, and vowel duration)

to identify vowels, the relative importance of the cues differed between groups. This was interpreted as an indication that “hearing loss alters the way in which acoustic cues are used to identify vowels” (p. 268), and furthermore, that the clear speech acoustic changes that lead to improved intelligibility for YNH listeners might not actually benefit listeners with hearing loss. For example, while the results of [Ferguson and Kewley-Port \(2007\)](#) suggested that vowel space expansion is an effective clear speech strategy for YNH listeners, this expansion did not improve vowel intelligibility for the EHI listeners in [Ferguson and Kewley-Port \(2002\)](#). In fact, raising F2 for front vowels was negatively associated with intelligibility, rendering them *less* intelligible in clear speech than in conversational speech.

On the other hand, the results of [Ferguson \(2012\)](#) suggest that the results of [Ferguson and Kewley-Port \(2002\)](#) may have been specific to their single talker. Of the 41 talkers in the Ferguson Database, only 3 showed the pattern of results shown by the 2002 talker (significant clear speech vowel intelligibility benefit for YNH listeners, no benefit for EHI listeners). Indeed, [Ferguson \(2012\)](#) found that the results for the two listener groups were strongly correlated ( $r = 0.75$ ), suggesting that the acoustic characteristics associated with improved vowel intelligibility for normal-hearing listeners may also be generally helpful for listeners with hearing loss. The previous literature on vowels in clear speech thus yields two competing predictions for the analyses carried out here:

- (1) The relationship between acoustic characteristics and vowel intelligibility in clear and conversational speech will *differ* between YNH and EHI listeners. This prediction is supported by the results of [Ferguson and Kewley-Port \(2002\)](#) as well as by studies showing differences between normal-hearing and hearing-impaired listeners for categorization of ambiguous vowel stimuli ([Molis and Leek, 2011](#)) and for formant discrimination ([Coughlin et al., 1998](#)). In particular, EHI listeners would be anticipated to rely more heavily on vowel duration than YNH listeners would. At the presentation level used in [Ferguson \(2012\)](#), the sloping high-frequency hearing losses of the EHI listeners would reduce the audibility of higher formants, which could make formant frequency information less important. In addition, the well-documented temporal processing deficits of older adults (e.g., [Kumar and Sangamanatha, 2011](#)) as well as age-related slowing in information processing (e.g., [Janse, 2009](#)) might make increased vowel duration an especially helpful clear speech acoustic change.
- (2) The relationship between acoustic characteristics and vowel intelligibility in clear and conversational speech will be the *same* for EHI listeners as it is for YNH listeners. This prediction is supported by the results of [Ferguson \(2012\)](#). It is also supported by the few signal processing experiments designed to determine the effects of specific clear speech acoustic changes on speech intelligibility that have included both normal-hearing and hearing-impaired listeners. For example, [Gordon-Salant](#)

(1986, 1987) found that increasing consonant-vowel intensity ratio improved consonant intelligibility for YNH, elderly normal-hearing, and EHI listeners, while increasing consonant duration did not. Krause and Braida (2009) also found similar results for listeners with and without hearing loss when speech was modified to increase either the energy in the 1000 to 3000 Hz range or modulation depth in the intensity envelope. Finally, neither normal-hearing nor hearing-impaired listeners have benefited when conversational speech has been slowed using signal processing (e.g., Uchanski *et al.*, 1996; Liu and Zeng, 2006).

The present study aims to investigate whether or not the relationships between acoustic measures and vowel intelligibility in conversational and clear speech differ between YNH and EHI listeners. Knowing the specific clear speech acoustic characteristics that result in improved speech understanding for listeners with hearing loss may help guide the development of new hearing aid algorithms as well as more specific clear speech training strategies for the communication partners of hearing-impaired individuals. For example, if reduced speaking rate were found to be more important for listeners with hearing loss than other clear speech acoustic changes, clear speech training could emphasize this slowing. Understanding the clear speech characteristics that make vowels more intelligible is an important step in the process of determining the clear speech characteristics that improve intelligibility in day-to-day communication.

## II. METHODS

### A. Materials

The materials consisted of 10 vowels (/i, ɪ, e, ε, æ, ɑ, ʌ, o, u, ʊ/) in /bVd/ context recorded in meaningful but neutral carrier sentences for the Ferguson Clear Speech Database (Ferguson, 2004). Forty-one talkers (21 female) aged 18–45 produced 7 tokens of each /bVd/ word, each centrally located in a different one of 16 possible carrier sentences. Example carrier sentences include, “Vera put the \_\_\_\_\_ on the table” and “I think the word \_\_\_\_\_ is hard for kids to say.” To elicit conversational speech, talkers were instructed to read the sentences aloud, speaking as they would in everyday conversation. To elicit clear speech, talkers were instructed to say the sentences as they would if they were talking to a person with hearing loss. Conversational speech was always recorded first, with the clear speech recording session taking place at least one day later. For each talker, two tokens of each vowel (usually the third and fourth out of the seven recorded) were selected from each speaking style for use in the perceptual experiments described in Ferguson (2004) and Ferguson (2012) as well as in the present analyses. Complete details regarding the talkers and the recording procedures may be found in Ferguson (2004).

### B. Intelligibility testing

Listeners’ vowel identification responses for the vowels described above were taken from the studies reported by Ferguson (2004) for YNH listeners and by Ferguson (2012)

for EHI listeners; full details about vowel identification testing are available in these papers. Briefly, the 7 YNH listeners were aged 19 to 30 yrs and had hearing thresholds  $\leq 20$  dB HL (ANSI, 2012) from 250 to 8000 Hz as determined by a pure-tone screening. The 40 EHI listeners were aged 65 to 87 yrs, and they had normal cognitive status and mild-to-moderately severe sloping sensorineural hearing losses with good word recognition abilities. Listeners were tested individually in a sound-treated room or booth, seated in front of a computer monitor, keyboard, and mouse. On each listening trial, one of the 1640 vowel stimuli was presented along with a segment of 12-talker babble. Listeners identified the vowel by selecting one of ten response categories shown on the monitor. Vowels were presented at 70 dB SPL for all listeners at SNRs of  $-10$  dB for the YNH listeners and  $-3$  dB for the EHI listeners. In Ferguson and Kewley-Port (2002) and in a pilot study, these SNRs had yielded a roughly comparable performance for vowels in conversational speech for the two groups while preventing ceiling effects for the most intelligible vowels. Test stimuli were divided into blocks of 100 to 120 items, with each block consisting of stimuli produced by 5 or 6 talkers of the same gender in the same speaking style. YNH listeners identified each stimulus three times, each time in a block composed of a different set of talkers, while EHI listeners identified each stimulus only once. About 1/3 of the EHI listeners heard different sets of 16 test blocks made up of different talker combinations to counter any effect that talker combination might have on the intelligibility of a given talker’s vowels.

### C. Acoustic analyses

Duration, steady-state formant frequencies, and dynamic formant movement were assessed. Vowel duration was computed for each vowel as the difference between onset and offset, which were determined manually using the waveform and spectrogram viewed in Cool Edit 2000. Steady-state and dynamic formant measures were derived from LPC formant tracks extracted using WaveSurfer (Sjölander and Beskow, 2006) set to a 20-ms Hamming window and a 10-ms frame rate. LPC order was normally  $M = 12$  but was adjusted as needed for each individual stimulus. Any tracking errors were corrected by hand editing. F1 and F2 values were extracted at several time locations specified by their distance from the vowel onset as a proportion of the vowel duration: 20%, 35%, 50%, 65%, and 80%. As in Ferguson and Kewley-Port (2002, 2007), vowel steady-state formant frequencies were extracted from the location 30 ms after the 20% point. F1 and F2 frequencies from the other landmarks were converted to Barks [using Eq. (6) from Traunmüller, 1990] and used to compute two dynamic metrics.

The first dynamic metric, vector length (VL), is the length of a vector in  $F1 \times F2$  space connecting the 20% and 80% values of the formants, calculated as a Euclidean distance:

$$VL = \sqrt{(F1_{80} - F1_{20})^2 + (F2_{80} - F2_{20})^2}. \quad (1)$$

Its use in previous studies of vowels in clear speech was motivated by pattern recognition studies showing substantial



gains in classifier performance when increasing the number of formant landmarks from one to two, but much smaller improvements when a third landmark was added (e.g., Hillenbrand *et al.*, 1995). Both VL and a mathematically similar measure, spectral change, were found to be significantly larger in clear speech than in conversational speech by Ferguson and Kewley-Port in 2002 and 2007, respectively. However, changes in dynamic formant movement were found to have a significant relationship with vowel intelligibility for only a few vowels in Ferguson and Kewley-Port (2002). Similarly, talkers who did and did not produce a clear speech vowel intelligibility benefit for YNH listeners showed comparable increases in the spectral change measure in Ferguson and Kewley-Port (2007).

The second dynamic metric, trajectory length (TL) is the sum of the lengths of four temporally equidistant vowel sections: From the 20% point to the 35% point, from 35% to 50%, from 50% to 65%, and from 65% to 80%. The length of each section (VSL) is computed as

$$\text{VSL}_n = \sqrt{(F1_n - F1_{n+1})^2 + (F2_n - F2_{n+1})^2}, \quad (2)$$

while TL is computed as

$$\text{TL} = \sum_{n=1}^4 \text{VSL}_n. \quad (3)$$

Fox and Jacewicz (2009) introduced this metric in a study comparing vowel production among different dialects of American English. They found that VL underestimated formant movement, particularly for vowels that have a U-shaped formant trajectory like /æ/ as produced by Southern speakers, and they devised TL to better capture the dynamic characteristics of such vowels.

## D. Statistical analyses

### 1. Statistical analysis of acoustic measures

Duration, steady-state F1 and F2 values, and the two dynamic metrics were analyzed using linear mixed-effects models (LMMs; Bates *et al.*, 2014) in R (R Core Team, 2013). This type of analysis allows for multiple random effects, which we use in the present study to model the random effects of individual talkers on acoustic measures (for reviews and examples of LMMs, see, e.g., Quené and Van den Bergh, 2004, 2008; Baayen, 2008; Baayen *et al.*, 2008; Quené, 2008). Fixed predictors in the LMMs were speaking style (dummy coding, 0 = conversational and 1 = clear), vowel (ten phonemes, converted to nine binary indicators), and the speaking style by vowel interaction. Random effects were the individual talkers as well as the individual differences between talkers in the effect of speaking style (Quené and Van den Bergh, 2004; Barr *et al.*, 2013). The LMM regression coefficients reported below may be regarded as population estimates of these fixed predictors, after correcting for the individual talkers' averages (intercepts) and for their individual differences in style effects (slopes). Consequently, these estimates may be regarded as more

reliable than those provided by conventional statistical techniques. Fixed estimates of the LMM were evaluated using their 95% confidence intervals (Kuznetsova *et al.*, 2013) against the null hypothesis that the fixed effect was zero. The random slope of the speaking style effect (i.e., the variability between talkers in the magnitude of the clear speech effect for each acoustic measure) was assessed by means of likelihood ratio tests, in which models with and without these random slope coefficients were compared by means of likelihood ratio tests (using restricted maximum likelihood estimation; Pinheiro and Bates, 2000). The LMM for each acoustic measure was based on two tokens for each combination of speaking style, vowel, and talker ( $N = 1640$ ).

### 2. Statistical analysis of intelligibility measures

Listeners' accuracy in vowel identification was modeled by means of Generalized Linear Mixed Modeling (GLMM; Quené and Van den Bergh, 2008; Bates *et al.*, 2014) in R. This technique is similar to logistic regression, where a binomial dependent variable (correct or incorrect response in the present analysis) is regressed onto multiple predictors (speaking style, listener group, duration, F1, F2, VL, and TL, described in more detail below). Like the LMMs used for the acoustic measures, the present GLMM includes random effects (intercepts) for individual talkers; these were crossed in the GLMM with random effects for individual listeners.

In the present GLMM, fixed predictors were (a) the categorical factors listener group (YNH ~ EHI), (b) speaking style, (c) binary contrast codes for three phonological features of the vowels, viz., binary contrasts "highness" (/i, ɪ, u, u/ coded as +1/2 = high and the other vowels as -1/2 = non-high), "backness" (/ɑ, ʌ, o, ʊ, u/ coded as +1/2 = back and the other vowels as -1/2 = non-back), and "tenseness" (/i, e, a, o, u/ were coded as +1/2 = tense and the other vowels as -1/2 = lax), and (d) the acoustic measures duration, F1, F2, VL, and TL. Some of these measures were transformed to obtain a more normal distribution, and all were centered to avoid multicollinearity (Baayen, 2008). In addition, measures VL and TL were by their nature highly correlated ( $r = 0.834$  across vowel tokens), so the predictor TL was adjusted by subtracting the VL component prior to analyses. The resulting measure  $\text{TL}_{\text{adj}}$  captures non-linear formant change, or the deviation of the trajectory from the straight-line vector VL (the correlation between VL and  $\text{TL}_{\text{adj}}$  was  $r = -0.209$  across vowel tokens). The transformations and center values for each acoustic measure are given in Table I. The predictors for listener group and for the phonological features used contrast codes (-1/2, +1/2) instead of dummy codes (0, 1), so that the estimate of a fixed predictor in the GLMM may be regarded as the average effect of that predictor across the two listener groups, and across phonological vowel classes. By using this contrast coding, the interactions describe the difference in regression slopes between the two groups or between opposite phonological classes.

Each fixed estimate of the GLMM was evaluated using its Z value (Quené and Van den Bergh, 2008; Hox, 2010)

TABLE I. Transformations performed and center values used for each acoustic measure prior to inclusion in the GLMM. VL = vector length; TL = trajectory length.

Acoustic measure	Transformation	Center value(s)
Duration (ms)	Square root	median = 15.5
F1 (Bark)	n/a	median = 5.297
F2 (Bark)	n/a	front median = 13.260 back median = 10.145
VL (Bark)	Square root	median = 1.062
TL <sub>adj</sub> (Bark)	Square root of TL <sub>adj</sub>	median = 0.641

against the null hypothesis that the fixed effect was zero. Individual talkers and listeners were added as random effects (intercepts). In addition, the acoustic measures were included in the random part of the GLMM at the talker level, i.e., the regression slopes for the acoustic measures were allowed to vary between talkers (“random slopes,” Snijders and Bosker, 1999; Quené and Van den Bergh, 2004; Barr *et al.*, 2013). More complex GLMMs were also attempted, but these could not be estimated reliably. Thus the random part of the optimal GLMM includes random slopes at the talker level for acoustic measures, but not for phonological features, and it does not contain any random slopes at the listener level. Again, the regression coefficients reported below may be regarded as population estimates of the fixed predictors, after correcting for the (random) individual talkers and individual listeners in this study, and after correcting for individual differences between talkers in these regression coefficients. The GLMM was based on 61 responses for each of the 1640 vowel tokens ( $N = 100\,040$  responses).

### III. RESULTS

#### A. Acoustic differences between clear and conversational speech

Mean values for vowel duration, VL, and TL<sub>adj</sub> are displayed for each talker in each speaking style in Table II. Clear/conversational ratios are also given for each metric for each talker.

##### 1. Duration

In general, talkers increased their vowel duration when speaking clearly, producing average vowel durations of 290 ms in clear speech and 204 ms in conversational speech. The LMM on the transformed durations confirmed that vowel lengthening in clear speech was significant for all vowels, as illustrated in Fig. 1. The individual differences between talkers (in their style effects, see Sec. II D) contributed significantly to the LMM ( $\chi^2 = 580.9$ ,  $df = 2$ ,  $p < 0.0001$ ).

##### 2. Steady-state measures

*a. F1.* Talkers produced vowels differing in F1 in conversational and in clear speech, as illustrated in Fig. 2, with vowels differing in the direction and significance of this change. For the high front vowels (/i/, /ɪ/) the effect is negative and not significant. For all other vowels except /u/, the

TABLE II. Average vowel duration, VL, and adjusted trajectory length (TL<sub>adj</sub>) values for 41 talkers in clear (CL) and conversational (CON) speech. The clear/conversational ratio (CL/CON) is also given for each metric.

Talker	Duration (ms)			VL (Barks)			TL <sub>adj</sub> (Barks)		
	CL	CON	CL/CON	CL	CON	CL/CON	CL	CON	CL/CON
F01	417	305	1.4	1.13	1.31	0.9	0.99	0.55	1.8
F02	227	174	1.3	1.32	1.28	1.0	0.50	0.75	0.7
F03	241	151	1.6	1.32	1.62	0.8	0.51	0.51	1.0
F04	210	194	1.1	1.23	1.23	1.0	0.51	0.52	1.0
F05	388	186	2.1	1.44	1.03	1.4	0.62	0.69	0.9
F06	288	173	1.7	1.53	1.39	1.1	0.63	0.51	1.2
F07	287	233	1.2	1.01	0.94	1.1	0.58	0.44	1.3
F08	362	187	1.9	1.44	1.49	1.0	0.94	0.54	1.7
F09	324	287	1.1	1.21	1.15	1.1	0.56	0.69	0.8
F10	282	193	1.5	1.32	0.88	1.5	0.64	0.55	1.2
F11	354	153	2.3	1.63	1.26	1.3	0.57	0.57	1.0
F12	276	180	1.5	1.63	1.32	1.2	0.78	0.73	1.1
F13	292	231	1.3	1.18	1.09	1.1	0.34	0.47	0.7
F14	246	184	1.3	1.51	1.28	1.2	0.87	0.88	1.0
F15	247	168	1.5	1.53	1.37	1.1	0.54	0.49	1.1
F16	301	195	1.5	1.55	1.16	1.3	0.51	0.44	1.2
F17	267	179	1.5	1.24	1.02	1.2	0.43	0.53	0.8
F18	282	211	1.3	1.47	1.32	1.1	0.58	0.48	1.2
F19	336	306	1.1	1.41	1.52	0.9	0.40	0.55	0.7
F20	330	221	1.5	1.47	1.51	1.0	0.52	0.51	1.0
F21	300	258	1.2	1.48	1.55	1.0	0.75	0.87	0.9
M01	304	187	1.6	1.32	1.04	1.3	0.46	0.41	1.1
M02	262	195	1.3	0.91	0.81	1.1	0.29	0.54	0.5
M03	247	128	1.9	1.17	1.04	1.1	0.27	0.49	0.6
M04	316	209	1.5	1.32	1.26	1.0	0.50	0.41	1.2
M05	252	155	1.6	0.79	1.18	0.7	0.39	0.45	0.9
M06	344	164	2.1	1.46	1.21	1.2	0.36	0.44	0.8
M07	309	209	1.5	1.13	0.95	1.2	0.63	0.44	1.4
M08	239	226	1.1	1.21	1.12	1.1	0.51	0.36	1.4
M09	297	176	1.7	1.14	1.06	1.1	0.24	0.24	1.0
M10	223	192	1.2	1.00	0.83	1.2	0.35	0.35	1.0
M11	336	194	1.7	1.02	0.88	1.2	0.47	0.38	1.2
M12	314	225	1.4	1.15	1.08	1.1	0.48	0.36	1.3
M13	321	216	1.5	1.65	1.80	0.9	0.40	0.44	0.9
M14	280	203	1.4	0.95	0.80	1.2	0.53	0.50	1.1
M15	255	232	1.1	0.88	1.04	0.8	0.43	0.34	1.3
M16	221	220	1.0	1.06	1.09	1.0	0.31	0.27	1.1
M17	231	220	1.1	1.23	1.11	1.1	0.40	0.53	0.8
M18	291	217	1.3	1.12	1.06	1.1	0.17	0.26	0.7
M19	345	288	1.2	1.06	1.00	1.1	0.20	0.20	1.0
M20	278	168	1.7	1.34	1.08	1.2	0.28	0.39	0.7
MEAN	291	205	1.5	1.27	1.1	0.09	0.50	0.49	1.0

effect is positive and significant ( $p < 0.05$ ). Notice that the LMM compares F1 values between styles within talkers, whereas the boxes in Fig. 2 summarize F1 values between styles across talkers. An across-talker difference in F1 may be found to be significant if within-talker differences are consistent (as was the case for /e/ in Fig. 2), while a similar-size across-talker difference in F1 may be found to be not significant if within-talker differences vary (as for /u/ in Fig. 2). All significant changes in F1 are in the positive direction, corresponding with a lower location (more open jaw and lower tongue) in the vowel space. The individual

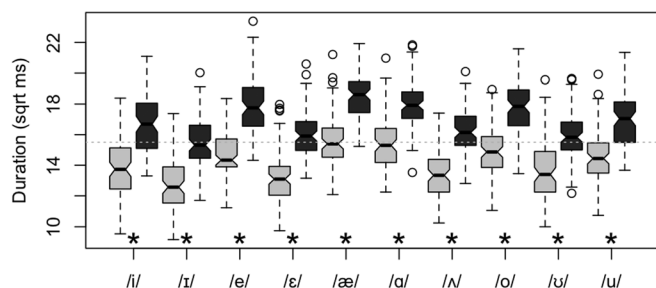


FIG. 1. Boxplot of square-root-transformed duration (in square root ms), broken down by vowel and by style (light boxes: Conversational, dark boxes: Clear). Boxes summarize the median (thick center line), lower and upper quartiles (lower and upper box limits), outlier fence values (lower and upper whiskers), and outliers (circles) of each data set. Notches in a box indicate the approximate 95% confidence intervals of the box' median. Asterisks indicate whether the clear speech change was significant ( $p < 0.05$ ) for that vowel, according to LMM. The horizontal reference line indicates the value to which the square-root-transformed duration was centered before data analyses.

differences between talkers contributed significantly to the LMM ( $\chi^2 = 18.8$ ,  $df = 2$ ,  $p < 0.0001$ ).

*b. F2.* Talkers' vowels also differed in F2 in conversational and in clear speech, as illustrated in Fig. 3, with vowels differing in the direction and significance of this change. For all five front vowels (left panel), the clear speech effect is positive and significant ( $p < 0.05$ ), corresponding with a more forward articulation of the front vowels in clear speech. However, for the back vowels (right panel) we see no change for /a/ or /ʌ/, and a significant negative change in F2 for the other three back vowels, corresponding with a more backward location of these vowels in the vowel space in clear speech. The results for F1 and F2 together indicate a more extreme position in the vowel space for several vowels: /e/, /ɛ/, and /æ/ are located lower and more frontward (higher F1, higher F2) in the vowel space in clear speech, whereas /o/ and /u/ are located also lower, but more backward (higher F1, lower F2) in the vowel space in clear speech. These vowel space shifts are illustrated in Fig. 4. The individual differences between talkers contributed significantly to the LMM ( $\chi^2 = 8.6$ ,  $df = 2$ ,  $p = 0.0133$ ).

### 3. Dynamic measures

VL differed between conversational and clear speech, but the direction and significance of the change varies between vowels, as illustrated in Fig. 5. For three front

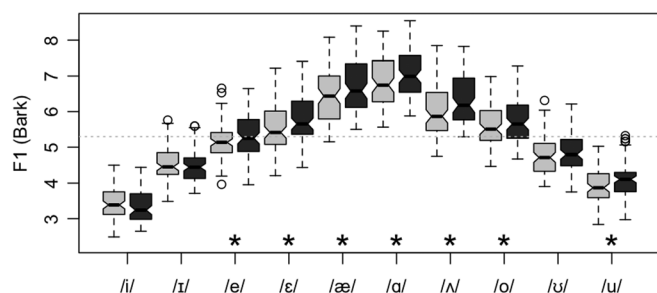


FIG. 2. Boxplot of F1 frequency (in Bark), broken down by vowel and by style (light boxes: Conversational, dark boxes: Clear; see Fig. 1 for symbolology). The horizontal reference line indicates the value to which the F1 frequency was centered before data analyses.

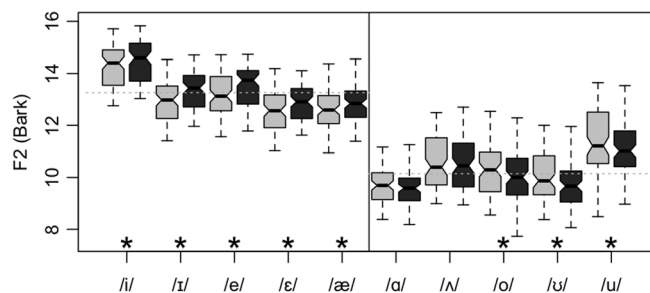


FIG. 3. Boxplot of F2 frequency (in Bark), broken down by vowel and by style (light boxes: Conversational, dark boxes: Clear; see Fig. 1 for symbolology). The horizontal reference lines indicate the values to which the F2 frequency was centered before data analyses; different values were used for front vowels (left panel) and for back vowels (right panel).

vowels, /i/, /ɛ/, and /æ/, the clear speech effect is negative and significant, corresponding with a smaller change (or more stable vowel location) in clear speech as compared to conversational speech. For the front vowel /e/, however, as well as for the three back vowels /o/, /u/, and /ʊ/, the clear speech effect is positive and significant, indicating for these vowels a larger change in F1 and F2 over the vowel nucleus, or more gliding, in clear speech as compared to conversational speech. VL did not differ significantly between clear and conversational speech for the vowels /ɪ/, /a/, and /ʌ/. The individual differences between talkers did not contribute to the LMM ( $\chi^2 = 1.4$ ,  $df = 2$ , n.s.).

$TL_{adj}$  also differed between conversational and clear speech, with effects varying between vowels, as illustrated in Fig. 6. For /e/ and /o/,  $TL_{adj}$  was smaller in clear speech as compared to conversational speech, indicating less deviation from a straight-line vector in the vowel space in clear speech. For /ɛ/ and /æ/, however, a positive effect was observed, indicating greater deviation from a straight-line trajectory, i.e., a more curved trajectory in clear speech for these vowels.  $TL_{adj}$  did not differ significantly between clear and conversational speech for /i/, /ɪ/, /a/, /u/, or /ʊ/. The individual differences between talkers contributed significantly to the LMM ( $\chi^2 = 9.1$ ,  $df = 2$ ,  $p = 0.0106$ ).

When both dynamic measures are taken together, we see different patterns for different vowels. Perhaps most clearly, for both /ɛ/ and /æ/ the VL decreases while  $TL_{adj}$

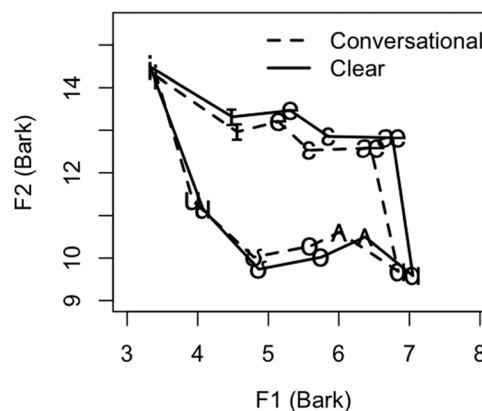


FIG. 4. Vowel spaces reflecting F1 and F2 frequencies (in Bark) for each vowel averaged across all talkers.

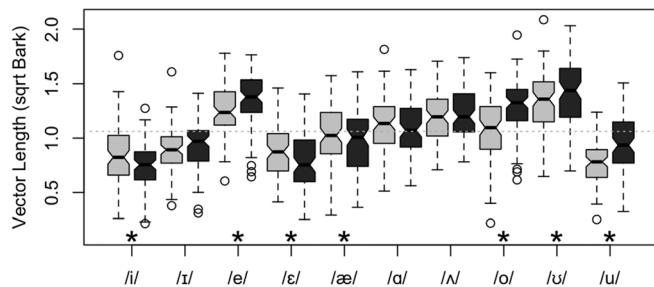


FIG. 5. Boxplot of square-root-transformed VL (in square root Bark), broken down by vowel and by style (light boxes: Conversational, dark boxes: Clear; see Fig. 1 for symbology). The horizontal reference line indicates the value to which the VL was centered before data analyses.

increases: For these vowels, straight-line distance traveled between the endpoints of the vowel nucleus is smaller in clear speech, but the movement that occurs within the vowel nucleus is greater. For /e/, /o/, and /u/, the opposite pattern is observed, with increased VL and decreased  $TL_{adj}$ : Overall straight-line formant frequency changes are larger, but there is less movement during the vowel nucleus in clear versus conversational speech. VL increases for /i/ and decreases for /u/ in clear speech, without any effect in  $TL_{adj}$  or curviness of the spectral trajectory. Finally, for /ɪ/, /a/, and /ʌ/ the formant dynamics are essentially the same in both styles, with no significant effects for either dynamic measure.

## B. Relationship between acoustic characteristics and intelligibility

The optimal GLMM for listeners' accuracy in vowel intelligibility is summarized in Table III. This GLMM contains fixed estimates for the main effects of (a) the listener group, (b) speaking style, (c) three phonological features, and (d) duration, steady-state formant frequencies, and dynamic measures; estimates were also included for the two-way interactions between tenseness and duration, between highness and F1, and between backness and F2. Furthermore, estimates were included for the interactions between the above main effects and listener group and between the above two-way interactions and listener group, and for the two-way interaction of listener group and

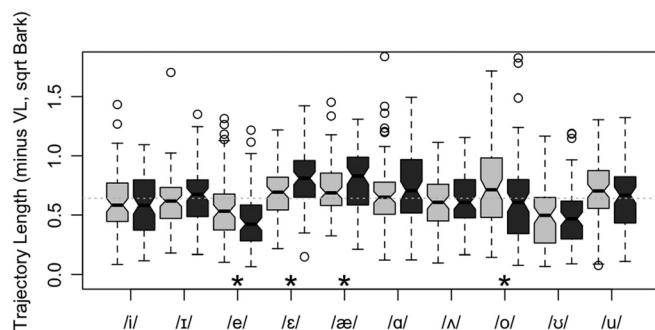


FIG. 6. Boxplot of adjusted and square-root-transformed trajectory length ( $TL_{adj}$ , in square root Bark), broken down by vowel and by style (light boxes: Conversational, dark boxes: Clear; see Fig. 1 for symbology). The horizontal reference line indicates the value to which  $TL_{adj}$  was centered before data analyses.

speaking style. Inclusion of other interactions, e.g., between tenseness and F1, did not improve the GLMM any further (according to likelihood ratio tests), and these interactions were therefore ignored.

The assessment of a GLMM is notoriously difficult (e.g., Gelman and Hill, 2007); here the optimal GLMM was compared against a null model containing only listener group and speaking style as fixed predictors, excluding phonological features and acoustic measures in its fixed part. The standardized residual errors of the optimal GLMM are indeed smaller than those of that null model, for nearly all talkers in both speaking styles, as illustrated in Fig. 7. The optimal model predicts the listeners' responses significantly better than the null model (likelihood ratio test,  $\chi^2 = 4711$ ,  $df = 22$ ,  $p < 0.0001$ ), although the relative reduction of standardized residual error (an evaluation measure somewhat comparable to the proportion of variance explained) is only small at 0.052 relative to the null model.

Each estimated GLMM regression coefficient ( $\beta$ ) is reported in Table III in log odds units, with a  $p$  value based on its standard error; each coefficient may also be interpreted as an odds ratio associated with that effect. First, we see that the accuracy or intelligibility differs significantly between the two listener groups. For the EHI listeners, the estimated overall accuracy was 73% [estimated log odds  $0.772 + (-1/2)(-0.491) = 1.018$ ], while for the YNH it was 63% [estimated log odds  $0.772 + (+1/2)(-0.491) = 0.527$ ;  $\beta = -0.491$ , odds ratio 1.63,  $p < 0.0001$ ]. This difference is not surprising; Ferguson (2012) also reported it and attributed it to SNR differences between the two groups during vowel identification testing.

Identification accuracy is far higher for tense than for lax vowels ( $\beta = 0.538$ , odds ratio 1.71,  $p < 0.0001$ ), and this difference is smaller for EHI listeners than for YNH listeners, as illustrated in the left panel of Fig. 8. In addition, vowel duration had a small but significant positive effect on the odds of a vowel being identified correctly, but only for tense vowels. The significant interaction between listener group, tenseness, and vowel duration corresponds to the finding that for EHI listeners, the regression slope is steeper (and positive) for tense than for lax vowels, whereas for YNH listeners the regression slopes of vowel duration on intelligibility are approximately equal (and zero) for tense and lax vowels (see Fig. 8, right panel). Conversely, the regression slope differs between the two listener groups for tense vowels, but not for lax vowels.

Identification accuracy is strongly correlated with F1. The significant interaction between the phonological feature highness and the acoustic measure F1 indicates that for high vowels the regression slope is negative [ $0.253 + (+1/2)(-1.254) = -0.374$ ], while for nonhigh vowels it is strongly positive [ $0.253 + (-1/2)(-1.254) = 0.880$ , see Fig. 9, left panel]. Moreover, the significant three-way interaction between listener group, highness, and F1 confirms that the interaction is stronger (larger difference in F1 slopes between high and nonhigh vowels) for the EHI listeners than for the YNH listeners (see Fig. 9, right panel); the difference between high and nonhigh vowels in direction of F1 slope is larger for EHI listeners than for YNH listeners.



TABLE III. Estimated parameters of the null model and of the optimal GLMM of vowel intelligibility. Estimates of fixed parameters are given with standard error (in parentheses) and  $Z$  and  $p$  values. Interaction effects are indicated using colons.  $N = 100\,040$ .

Fixed effects	Null model				Optimal model			
	Estimate	s.e.	$Z$	$p$	Estimate	s.e.	$Z$	$p$
Intercept, conversational style	1.452	(0.124)	11.69	<0.0001	0.772	(0.128)	6.01	<0.0001
Group <sup>b</sup>	−0.714	(0.130)	−5.48	<0.0001	−0.491	(0.138)	−3.55	<0.0001
Style <sup>a</sup>	0.107	(0.023)	4.61	<0.0001	0.293	(0.026)	11.40	<0.0001
Duration <sup>c</sup>					0.012	(0.022)	0.53	n.s.
Tenseness <sup>b</sup>					0.537	(0.025)	21.48	<0.0001
F1 <sup>c</sup>					0.253	(0.056)	4.50	<0.0001
Highness <sup>b</sup>					−0.059	(0.038)	−1.55	0.1201
F2 <sup>c</sup>					0.257	(0.056)	4.60	<0.0001
Backness <sup>b</sup>					−0.769	(0.024)	−32.30	<0.0001
VL <sup>c</sup>					0.576	(0.113)	5.10	<0.0001
TL <sub>adj</sub> <sup>c</sup>					−0.032	(0.134)	−0.24	n.s.
Group: Style	−0.025	(0.033)	−0.79	n.s.	−0.025	(0.043)	−0.59	n.s.
Group: Duration					−0.009	(0.010)	−0.95	n.s.
Group: Tenseness					0.510	(0.044)	11.58	<0.0001
Duration: Tenseness					0.037	(0.009)	4.41	<0.0001
Group: F1					0.273	(0.028)	9.72	<0.0001
Group: Highness					0.868	(0.062)	14.04	<0.0001
F1: Highness					−1.254	(0.031)	−40.12	<0.0001
Group: F2 <sub>adj</sub>					0.293	(0.020)	14.72	<0.0001
Group: Backness					−0.736	(0.039)	−19.00	<0.0001
F2 <sub>adj</sub> : Backness					−0.114	(0.023)	−4.88	<0.0001
Group: VL					0.117	(0.064)	1.82	0.0684
Group: TL <sub>adj</sub>					−0.057	(0.070)	−0.82	n.s.
Group: Duration: Tenseness					−0.034	(0.015)	−2.18	0.0291
Group: F1: Highness					0.294	(0.049)	5.96	<0.0001
Group: F2 <sub>adj</sub> : Backness					0.032	(0.037)	0.87	n.s.
Random effects	variance				variance			
Intercept   Talkers	0.4941				0.4678			
Duration   Talkers	0.0388				0.0191			
F1   Talkers	0.0925				0.1111			
F2   Talkers	0.2803				0.1159			
VL   Talkers	0.8561				0.4523			
TL <sub>adj</sub>   Talkers	1.0121				0.7668			
Intercept   Listeners	0.2267				0.2451			

<sup>a</sup>Dummy coding (0, 1)

<sup>b</sup>Contrast coding (−1/2, +1/2)

<sup>c</sup>Transformed and centered to values reported in Table I.

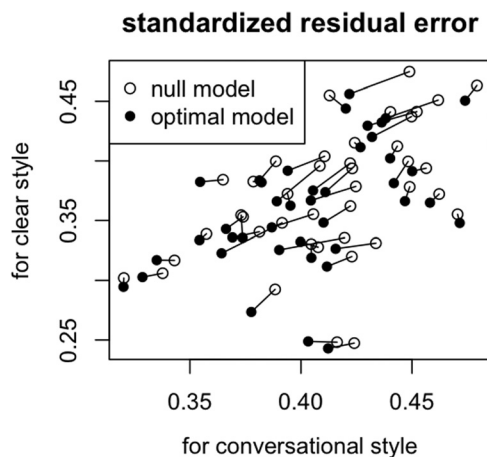


FIG. 7. Standardized residual errors of the optimal GLMM (filled symbols) and the null model (open symbols), broken down by talkers and by speaking styles (see text).

Similarly, identification accuracy is also correlated with F2. For YNH listeners the regression slopes are positive, both for back and nonback vowels. For EHI listeners, the regression slopes are shallower: Approximately zero for back vowels, and moderately positive for nonback vowels (see Fig. 10, left panel). The significant interaction between the phonological feature backness and the acoustic measure F2 indicates that for back vowels the regression slope is slightly positive [ $0.257 + (+1/2)(-0.114) = 0.200$ ], while for nonback vowels the regression slope is significantly more positive [ $0.257 + (-1/2)(-0.114) = 0.314$ , see Fig. 10, left panel]. The difference in slopes between back and nonback vowels is approximately the same for both listener groups (see Fig. 10, right panel), as indicated by the absence of a significant interaction between F2, group, and backness.

With regard to the dynamic spectral measures, the regression analysis results in significantly positive slopes for



### effect of duration (sqrt ms, centered)

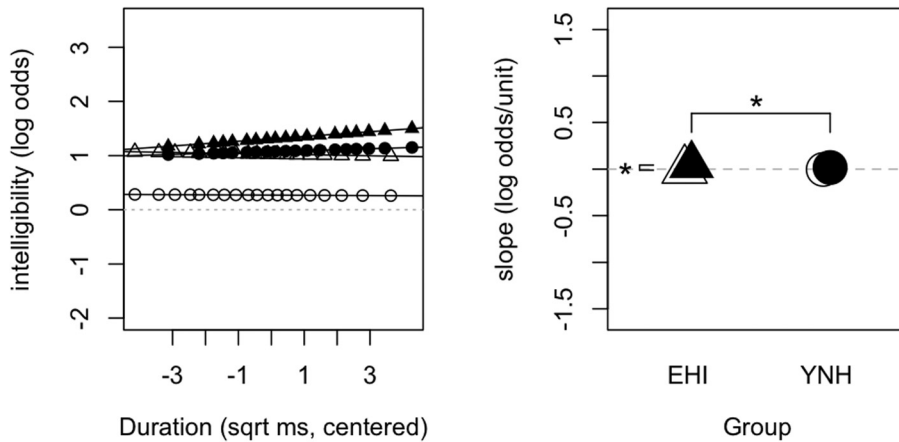


FIG. 8. Estimated regression lines (left panel) and regression slopes (right panel) of identification accuracy (in log odds) on vowel duration (in square root ms, centered), broken down for tense vowels (filled symbols) and lax vowels (open symbols), and for EHI (triangles) and YNH listeners (circles). In the left panel, identifier symbols are spaced at 0.05 quantiles of the predictor (from 0.05 to 0.95 quantiles). In the right panel, the asterisk indicates that the regression slopes for tense vowels differ significantly ( $p_{\text{MCMC}} < 0.05$ ) between listener groups.

### effect of F1 (Bark, centered)

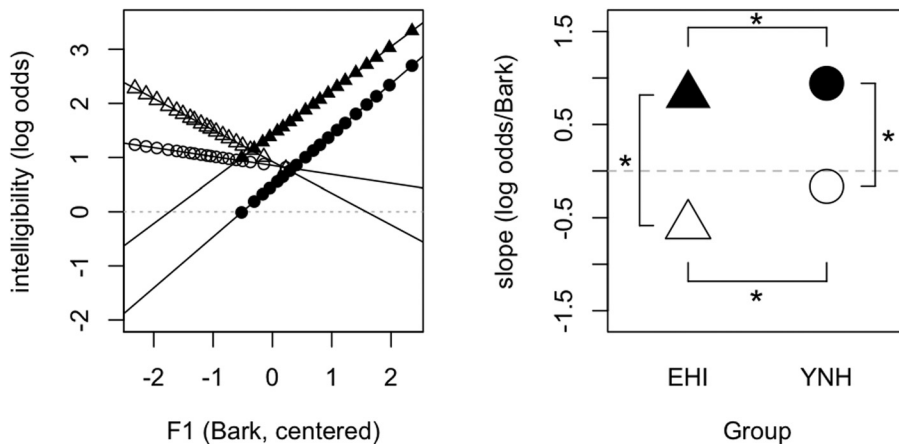


FIG. 9. Estimated regression lines (left panel) and regression slopes (right panel) of identification accuracy (in log odds) on F1 frequency (in Bark, centered), broken down for high vowels (open symbols) and nonhigh vowels (filled symbols), and for EHI (triangles) and YNH listeners (circles). In the left panel, identifier symbols are spaced at 0.05 quantiles of the predictor (from 0.05 to 0.95 quantiles). In the right panel, the asterisk indicates that the regression slopes for high vowels differ significantly ( $p_{\text{MCMC}} < 0.05$ ) between listener groups.

### effect of F2 (Bark, centered)

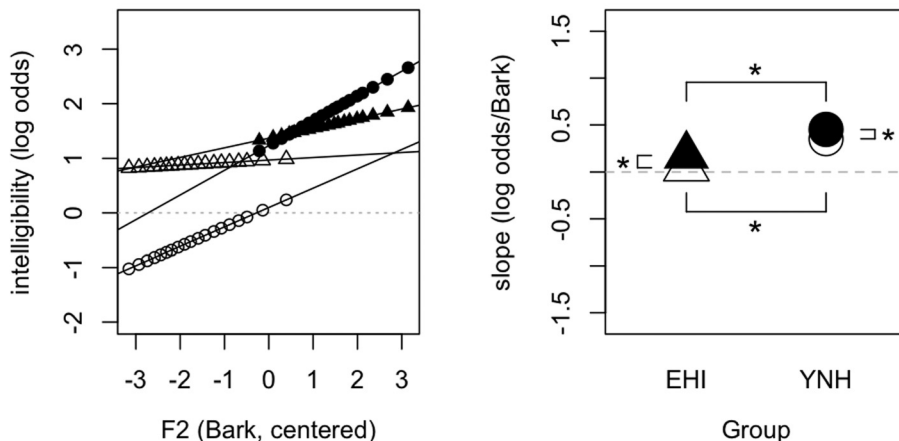


FIG. 10. Estimated regression lines (left panel) and regression slopes (right panel) of identification accuracy (in log odds) on F2 frequency (in Bark, centered), broken down for back vowels (open symbols) and front vowels (filled symbols), and for EHI (triangles) and YNH listeners (circles). In the left panel, identifier symbols are spaced at 0.05 quantiles of the predictor (from 0.05 to 0.95 quantiles). In the right panel, asterisks indicate that the regression slopes for back and front vowels differed significantly ( $p_{\text{MCMC}} < 0.05$ ) between listener groups.

## effect of Vector Length (sqrt Bark, centered)

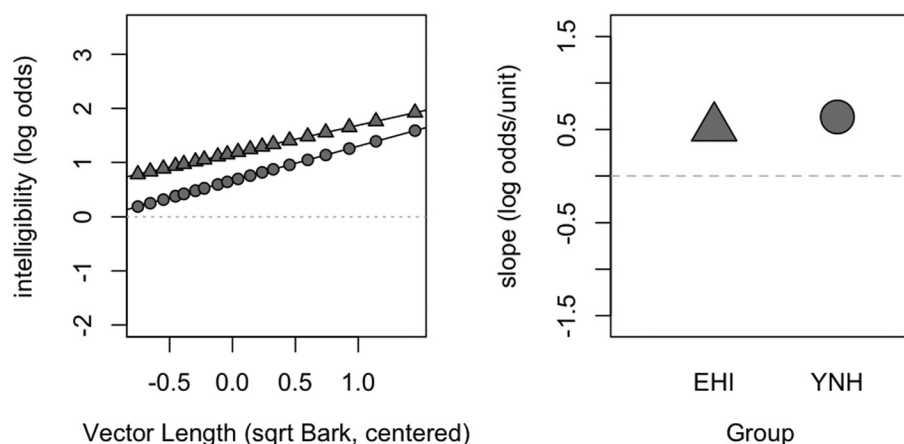


FIG. 11. Estimated regression lines (left panel) and regression slopes (right panel) of identification accuracy (in log odds) on VL (in square root Bark, centered), broken down for EHI (triangles) and YNH listeners (circles). In the left panel, identifier symbols are spaced at 0.05 quantiles of the predictor (from 0.05 to 0.95 quantiles).

formant VL (see Fig. 11): vowels with a larger net change in the F1-by-F2 vowel space have better odds of being identified correctly. This effect tends to be somewhat larger for the YNH group as indicated by the marginally significant interaction effect between VL and listener group. The other dynamic measure,  $TL_{adj}$  (see Sec. II D 2), has a small but non-significant negative effect on the accuracy of vowel identification: having a more curved (non-linear) trajectory in the vowel space does not affect the odds of a vowel being identified correctly (Fig. 12) for either listener group.

The regression coefficients show that even when all the phonological features and acoustic measures are taken into account, intelligibility remains higher for clear vowels than for conversational vowels: the log-odds of a correct response increase by  $\beta = 0.293$  (odds ratio 1.34,  $p < 0.0001$ ) in clear versus conversational speech. (The null model containing only listener group and speaking style as fixed predictors, excluding phonological features and acoustic measures, yielded a smaller style effect of  $\beta = 0.107$ , odds ratio 1.11, cf. Table III, but this was offset by far larger variances in its random part; an even simpler null model, with random intercepts but without random slopes, yielded  $\beta = 0.243$  for the style effect, odds ratio 1.28). This suggests that the

perceptual benefit of clear speech is not explained entirely by the acoustic measures and phonological features (and their interactions) reported above. The two-way interaction between style and listener group was not significant, which indicates that the remaining clear speech *benefit* in vowel intelligibility is approximately the same for EHI listeners [benefit is  $0.293 + (-1/2)(-0.025) = 0.306$  logit units, odds ratio 1.36] and for YNH listeners [benefit is  $0.293 + (+1/2)(-0.025) = 0.281$  logit units, odds ratio 1.32] after the phonological features and the acoustic measures of the stimulus vowels (and their interactions) have been taken into account.

Finally, the GLMM with random slopes allows us to assess the correlations between talkers' random intercepts and slopes for each of the acoustic measures (see Table IV). Focusing on correlations with absolute values  $> 0.3$ , the correlations between these random coefficients suggest that talkers who have an above-average positive slope for duration tend to have a below-average slope for F1 and for F2. Talkers' slopes for F1 and F2 are positively correlated, as are the talkers' slopes for VL and  $TL_{adj}$ . This suggests that if a given talker's intelligibility depends more on vowel duration, then it depends less on static formant values, and vice versa. If a talker's intelligibility depends relatively strongly

## effect of Trajectory Length (adjusted, sqrt Bark, centered)

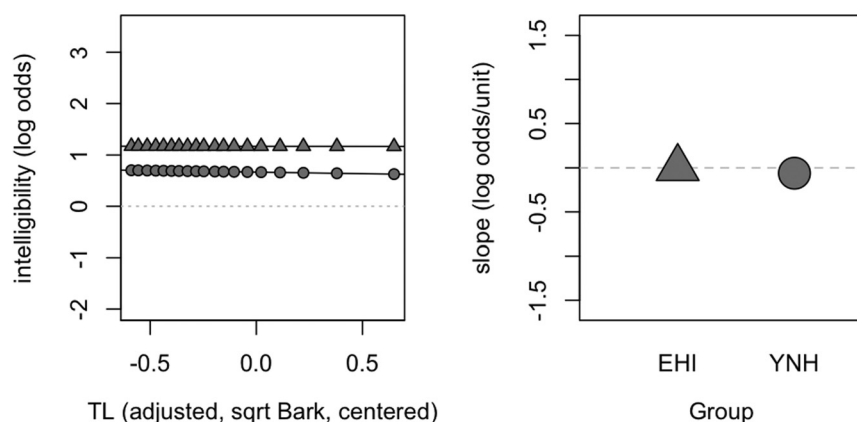


FIG. 12. Estimated regression lines (left panel) and regression slopes (right panel) of identification accuracy (in log odds) on adjusted and square-root-transformed trajectory length ( $TL_{adj}$ , in square root Bark, centered), broken down for EHI (triangles) and YNH listeners (circles). In the left panel, identifier symbols are spaced at 0.05 quantiles of the predictor (from 0.05 to 0.95 quantiles).

TABLE IV. Correlations between random coefficients (random intercept of talkers and random slopes of acoustic measures) at the talkers level of the optimal GLMM. Predictors were transformed and centered to values as reported in Table I.

	Intercept	Duration	F1	F2	VL	TL <sub>adj</sub>
Intercept	1					
Duration	0.108	1				
F1	0.229	-0.409	1			
F2	-0.027	-0.339	0.543	1		
VL	0.024	0.047	0.144	0.427	1	
TL <sub>adj</sub>	0.293	0.236	0.023	0.087	0.609	1

on static F1, then it also depends relatively strongly on static F2, and likewise for VL and TL<sub>adj</sub>.

## IV. DISCUSSION

### A. Acoustic differences between clear and conversational speech

#### 1. Duration

Averaged across all 41 talkers, vowel duration increased in clear speech by a factor of 1.5. This was expected based on previous data from the Ferguson Database as well as on other studies of clear speech. Ferguson and Kewley-Port (2007) reported vowel acoustic data for 10 of the 41 talkers and reported a clear/conversational vowel duration ratio of 1.4. In addition, Ferguson *et al.* (2010) measured speaking rate for all 41 talkers using sentence materials; the mean speaking rate in clear speech was 111 words per minute, significantly slower than the conversational speaking rate of 155 words per minute. Studies using other recorded materials have also reported longer vowel durations in clear speech. Clear/conversational vowel duration ratios computed using data reported in tables and figures were approximately 1.3 in Picheny *et al.* (1986), Moon and Lindblom (1994), and Lam *et al.* (2012), while Ferguson and Kewley-Port (2002) reported a duration ratio of 2.1 for their single talker.

As seen in Table II and confirmed in the LMM, variability among the talkers was extensive, with clear/conversational duration ratios ranging from 1.0 (talker M16) to 2.4 (talker F11). However, no talkers showed ratios <1. Most (34/41) talkers slowed down moderately (ratios between 1.00 and 1.69) while nine, or one-fifth, of the talkers showed duration ratios >1.7 (four of whom showed ratios >2), indicating substantial vowel duration increases (i.e., slowing down) by 69% or more in their clear speech. Few previous studies report individual talker data with regard to clear speech acoustic characteristics, but the talker variability gleaned from their figures and tables is consistent with the range observed here. Clear/conversational vowel duration ratios computed from these papers ranged from 1.1 to 1.7 among the five talkers in Moon and Lindblom (1994), while each of the three talkers in Picheny *et al.* (1986) showed a ratio of about 1.3. The ratio between speaking rates in habitual and clear speech in Rosen *et al.* (2011) ranged from 1.0 to 2.1 among the ten talkers, while the five English-speaking

talkers in Smiljanic and Bradlow (2005) showed clear/conversational speaking rate ratios ranging from 1.2 to 1.6. This suggests that when asked to speak clearly, nearly all talkers slow down, but generally by a factor of less than 2.

### 2. Steady-state measures

*a. F1.* When evaluated across all 41 talkers, F1 increased significantly in clear speech for 7 of the 10 vowels (i.e., all except /i/, /ɪ/, and /u/). Ferguson and Kewley-Port (2002) observed this in their single talker, noting that such shifts were associated with increased vocal effort in a study by Liénard and DiBenedetto (1999), and Ferguson and Kewley-Port (2007) reported it in their analysis of 10 of the 41 talkers. Increased F1 frequencies are also associated with a lower tongue position in the mouth, which would be likely to occur if talkers speak more loudly when speaking clearly. That is, talkers may open their mouths wider in order to become louder, resulting in a lower tongue position for all vowels. Ferguson *et al.* (2010) reported that on average, the present talkers spoke with 4 dB greater intensity in clear speech than conversational speech. The association between increased vocal intensity and increased F1 was explored by correlating unpublished speech intensity data from Ferguson (2002) with the average F1 change each talker produced in clear speech. For each talker, this was computed by determining the clear-conversational difference for each vowel, and then averaging across the 10 vowels. A strong positive relationship ( $r=0.69$ ) was observed between the level change (in dB) between conversational and clear speech and the concomitant change in F1 (in Bark). That is, talkers who showed greater speech intensity increases when speaking clearly also generally showed larger increases in F1.

F1 increases have occasionally appeared in other clear speech studies, such as Picheny *et al.* (1986). In their Fig. 8, talker MP shows a general upward shift on the F1 dimension for most of the vowels analyzed, while the other two talkers show this for just a few vowels. Krause and Braida (2004) also reported increases in F1 for some vowels when comparing clear speech spoken at a normal rate to conversational speech. However, most previous clear speech studies have focused on the vowel space rather than individual formant frequencies. Vowel space is usually computed using both F1 and F2, but a few papers have examined F1 and F2 separately. F1 range expansions of various magnitudes have been reported by Bradlow *et al.* (2003), by Hazan and Baker (2011), and by Lam *et al.* (2012). F1 range was computed for the talkers in the present study by subtracting the average F1 (in Bark) for the low vowels /æ/ and /a/ from the average F1 for the high vowels /i/ and /u/.

The clear/conversational ratio for F1 range was 1.09 on average [95% C.I. (1.05, 1.13),  $t(40)=4.3$ ,  $p=0.0001$ ; the 9% expansion corresponds with an average increase of the F1 range by 0.23 Bark), with large individual differences. Nine talkers showed ratios <1, indicating compressed (not expanded) F1 ranges in their clear speech, 25 talkers showed a moderate expansion (ratio between 1.00 and 1.21), and nine, or one-fifth, of the talkers showed ratios >1.18, indicating substantial expansion of their F1 range by 18% or

more in their clear speech. Similarly, Lam *et al.* (2012) reported that 11 of their 12 talkers showed trends toward increased F1 range in their various clear speech conditions. The pattern of significant F1 increases across vowels in the present data is in fact consistent with this expansion of the F1 range; the three vowels for which F1 did not increase were all high vowels. Thus it appears that when speaking clearly, most talkers raise F1 by lowering their jaw and tongue more than in conversational speech, but they do so in a selective manner that increases the acoustic distance between high and low vowels.

*b. F2.* While the dominant trend in the F1 data was an overall increase in F1, for F2 the direction of change varied among vowel categories. Specifically, the front vowels /i/, /I/, /e/, /ɛ/, and /æ/ showed significantly higher F2 values in clear speech, while the back vowels /o/, /ʊ/, and /u/ had significantly lower F2 values (and /ɑ/ and /ʌ/ were unchanged). A very similar pattern was observed for the single talker in Ferguson and Kewley-Port (2002), with significant positive shifts for all front vowels and smaller negative shifts for only two of the back vowels (/ʌ/ and /ʊ/). Ferguson and Kewley-Port's (2007) analysis of 10 of the 41 present talkers also revealed differential F2 behavior in clear speech for front versus back vowels, showing significant increases for the former and no significant change for the latter.

Combining increased front vowel F2 with decreased back vowel F2 yields an expanded F2 range, consistent with the overall vowel space expansion widely reported in the clear speech literature as well as with the F2 range expansions reported by Bradlow *et al.* (2003), Hazan and Baker (2011), and Lam *et al.* (2012). F2 range was computed here by subtracting the average F2 for the back vowels /ɑ/ and /u/ from the average F2 for the front vowels /i/ and /æ/. The clear/conversational ratio for F2 range was 1.14 on average [95% C.I. (1.09, 1.19),  $t(40) = 5.6$ ,  $p < 0.0001$ ; the 14% expansion corresponds with an average F2 range increase of 0.38 Bark], again with large individual differences. Eight talkers showed a ratio  $< 1$ , indicating a compressed (not expanded) F2 range in their clear speech, 25 talkers showed a moderate expansion (ratios between 1.00 and 1.27), and eight, or one-fifth, of the talkers showed ratios  $> 1.22$ , indicating substantial expansion of their F2 range by 22% or more in their clear speech. This suggests that raising F2 for front vowels (i.e., a more fronted articulation) and either maintaining or lowering it for back vowels (i.e., a more extreme back articulation) is a very common clear speech strategy.

### 3. Dynamic measures

The direction and magnitude of the clear speech effect for dynamic formant movement depended not only on the vowel but also on the specific dynamic metric. Three vowels, /I/, /ɑ/, and /ʌ/, showed no change on either metric. That is, for these vowels, neither the length of a linear trajectory linking the endpoints of the vowel nucleus (VL) nor the amount of non-linear deviation from that trajectory (TL<sub>adj</sub>) changed significantly between conversational and clear speech. The back vowels /ʊ/ and /u/ showed significant increases in VL with no change in TL<sub>adj</sub>. That is, the distance

traveled over the F1  $\times$  F2 space from the 20% point to the 80% point increased but the "curviness" that occurred on top of this overall movement did not change. Two other vowels, /e/ and /o/, also showed significant increases in VL in clear speech, accompanied with significant decreases in TL<sub>adj</sub>; the formant trajectories of these vowels became longer, but also less curved. Finally, the vowels /i/, /ɛ/, and /æ/ had smaller VL values in clear speech than in conversational speech; this was accompanied by a significant increase in TL<sub>adj</sub> for /ɛ/ and /æ/ but no change in TL<sub>adj</sub> for /i/. That is, the formant trajectories became shorter for /i/, /ɛ/, and /æ/, but also more curved for /ɛ/ and /æ/. This complicated, vowel-dependent relationship between VL and TL<sub>adj</sub> in clear and conversational speech confirms that the two metrics capture very different aspects of dynamic formant movement.

To date, only three studies have explored vowel formant dynamics in clear speech. All included a measure equivalent to VL, but only one reported individual vowel data. For the single talker in Ferguson and Kewley-Port (2002), VL increased significantly in clear speech for seven vowels: /i/, /I/, /e/, /ɛ/, /o/, /ʊ/, and /u/. These results are consistent with the present data showing increased formant movement for the highly dynamic /e/ and /o/ and for the back vowels /ʊ/ and /u/ and no change for the central vowels /ɑ/ and /ʌ/, but disagree for the front vowels /i/, /I/, and /ɛ/. Ferguson and Kewley-Port (2007) reported a metric they called "spectral change" and designated as  $\lambda$  for a subset of the present talkers, but only reported an average  $\lambda$  value taken across the five vowels that had the largest  $\lambda$  values in conversational speech (/i/, /ɑ/, /ʌ/, /o/, and /ʊ/). While this average  $\lambda$  measure was significantly greater in clear speech, the magnitude of the increase varied widely among the ten talkers described. Finally, Lam *et al.* (2012) reported significant increases in Ferguson and Kewley-Port's  $\lambda$  metric for both the set of tense vowels and the set of lax vowels for a group of 12 talkers. Of the three types of clear speech instructions compared by Lam *et al.*, the instructions that most closely resembled those given to the present study (which they called the "hearing impaired" instructions) yielded the largest increases in  $\lambda$ .

The across-vowel average data shown in Table II reveal considerable talker variability for both dynamic formant movement measures. Averaged across vowels, the clear/conversational ratio for VL was 1.09 on average [95% C.I. (1.04, 1.14),  $t(40) = 3.8$ ,  $p = 0.0005$ ]. Ten talkers showed a ratio  $< 1$ , indicating shorter (not longer) formant trajectories in their clear speech, 23 talkers showed a moderate increase (ratios between 1.00 and 1.21), and 8 or one-fifth of the talkers showed ratios  $> 1.21$ , indicating substantial increase of their VL by 21% or more in their clear speech. For TL<sub>adj</sub> or curviness of the formant trajectory, the clear/conversational ratio was 1.03 on average [95% C.I. (0.94, 1.12),  $t(40) < 1$ , n.s.]. Almost half of the talkers (19/41) showed a ratio  $< 1$ , indicating less nonlinearity (less curviness) of formant trajectories in their clear speech, 13 talkers showed a moderate increase (ratios between 1.00 and 1.23), and 9 or one-fifth of the talkers showed higher ratios  $> 1.23$ .

Surprisingly, while individual talker differences in the clear speech effect were statistically significant in the LMMs



for all of the other acoustic metrics, they were not significant for VL. Individual vowel data offer a possible explanation for this. Consider the five vowels for which both dynamic metrics changed significantly (/ɛ/ and /æ/, for which VL decreased and  $TL_{adj}$  increased, and /e/, /o/, and /u/, for which VL increased and  $TL_{adj}$  decreased). Individual talker data for each vowel were examined to determine how many talkers made clear speech changes consistent with the overall pattern, using an arbitrary change criterion of 1.05 (5% change) for both dynamic measures. Only for /o/ did the majority of talkers show the pattern of change observed in the LMM; 26 talkers showed both increased VL and decreased  $TL_{adj}$ . For /e/ and /u/, 17 and 16 talkers, respectively, increased VL and decreased  $TL_{adj}$ ; for /ɛ/ and /æ/, the pattern of decreased VL and increased  $TL_{adj}$  was shown by 19 and 16 talkers, respectively. Only 2 talkers matched the overall pattern for all five of these vowels; 12 matched one or fewer. We thus conclude that clear speech changes in vowel formant movement depend not only on the vowel and on the specific dynamic metric but also strongly depend on the talker.

## B. Relationship between acoustic characteristics and intelligibility

### 1. Duration and steady-state formant information

Two competing predictions were offered in Sec. I regarding the relationship between acoustic characteristics and vowel intelligibility in clear and conversational speech for YNH versus EHI listeners. The first stated that the relationship would differ for the two listener groups, while the second stated that the relationship would be the same. The results of the GLMM supported the first prediction: duration and F1 had a stronger effect and F2 had a weaker effect on vowel intelligibility for EHI listeners than for YNH listeners. Longer duration was associated with better intelligibility, but only for EHI listeners identifying tense vowels. For high vowels, lower F1 values were associated with higher intelligibility for EHI listeners while F1 had no effect on intelligibility for YNH listeners. Higher F2 values were associated with better intelligibility for both front and back vowels for YNH listeners, while F2 played a much weaker role for front vowels and was not related to intelligibility for back vowels for EHI listeners. The role of dynamic formant movement was the same for the two groups although YNH listeners tended to be somewhat more sensitive to VL (linear change).

The first prediction was based on an assumption that EHI listeners' vowel perception would be affected by their high-frequency hearing loss and by their age. Their hearing loss was expected to make vowel formants less audible, while their age was expected to cause both poorer temporal processing and slower cognitive processing. Taken together, these deficits could cause EHI listeners to rely more on vowel duration and less on spectral information than YNH listeners. The results for duration match this account, at least for tense vowels. Longer vowel duration helped both groups but especially EHI listeners, perhaps by allowing more processing time for vowel identifications and/or by increasing the temporal contrast between spectrally similar tense-lax pairs

like /i/-/ɪ/ and /u/-/ʊ/. Likewise, the results for F2 seem to support the idea that sloping hearing loss made higher-frequency formant information less available or less reliable as a cue to vowel identity for the EHI listeners. However, F2 was unrelated to intelligibility for EHI listeners only for back vowels, for which F2 ranged from 1100 to 1600 Hz. For front vowels, with F2 values between 1800 and 2500 Hz, higher F2 values were associated with higher intelligibility despite declining hearing thresholds in this range, albeit less strongly than was the case for YNH listeners. The F1 results are also difficult to explain in terms of audibility, as group differences occurred only for high vowels, the F1 values of which (350 to 500 Hz) fall in a range in which the EHI listeners had normal hearing. EHI listeners performed better when high vowels had lower F1 values, while F1 had no effect on intelligibility of high vowels for YNH listeners. For low vowels, whose F1 values span the 550 to 750 Hz range, higher F1 values were associated with better intelligibility to the same degree for both listener groups.

Thus, EHI listeners did use F1 and F2 information as well as vowel duration when identifying vowels, but did so differently from YNH listeners. The current results are therefore broadly consistent with those of [Ferguson and Kewley-Port \(2002\)](#) despite major methodological differences (many tokens per vowel and style for a single talker versus just two tokens per vowel and style for 41 talkers; within-vowel stepwise regression versus across-vowel mixed-effects models; very different intelligibility results for EHI listeners). Although, as [Ferguson and Kewley-Port \(2002\)](#) concluded, these results suggest that hearing loss alters the way acoustic cues are used for identifying vowels, the present results are not satisfactorily explained by group differences in hearing sensitivity. Part of the difficulty may lie in the fact that the EHI listeners heard the vowel stimuli at a much more favorable SNR than the YNH listeners (−3 vs −10 dB), meaning that audibility was driven by more than just the audiogram. Analyses including conditions where YNH and EHI listeners identify vowels at the same SNR are necessary in order to definitively determine whether and how these groups differ in the acoustic cues that underlie vowel intelligibility in clear and conversational speech.

The correlations among random coefficients at the talker level suggest that the contributions of acoustic measures to vowel intelligibility differ across individual talkers. For talkers for whom the contribution of duration to vowel intelligibility is relatively strong (above-average slope of duration), the contributions of F1 and of F2 are relatively weak (below-average slopes of F1 and of F2). In other words, talkers differ in the perceptual effectiveness of the acoustic cues they use to convey vowel identity; if duration contributes relatively much to vowel intelligibility then static formant values contribute relatively little, and vice versa. The strongest correlation is between the talkers' slopes for the two dynamic measures; when VL contributes relatively much, so does  $TL_{adj}$ , and vice versa. The pattern of correlations in Table IV suggests that there are three distinct factors (clusters of cues) that contribute to a talker's intelligibility: duration, static formant values, and dynamic formant measures. Cues within each cluster are positively correlated, whereas

cues between these clusters are either uncorrelated or negatively correlated.

## 2. Dynamic formant movement

For both listener groups (slightly more so for the YNH listeners), greater linear formant movement between the endpoints of the vowel nucleus (VL) was associated with higher vowel intelligibility, while non-linear movement within the nucleus ( $TL_{adj}$ ) was unrelated to vowel intelligibility. The fact that dynamic formant movement played a significant role in intelligibility for both groups is consistent with the results of Ferguson and Kewley-Port (2002), although Ferguson and Kewley-Port (2007) found no differences in formant dynamics for talkers who did and did not produce a clear speech vowel intelligibility benefit for YNH listeners. In general, the present line of research has assumed that increased formant movement in clear speech would make vowels easier to identify, particularly in challenging situations where listeners might use moving formants to help them segregate speech from competing multi-talker babble. However, only linear movement was helpful for intelligibility. Larger amounts of nonlinear movement within the nucleus, in contrast, did not affect vowel intelligibility.

This latter result is consistent with research suggesting that the details of formant movement during the vowel nucleus contribute little to vowel identification. Neel (2004) found improved listener performance when vowels were synthesized using formant values sampled from two or three locations in the vowel versus only one, but no additional improvement from finer sampling. Pattern classifier performance also improves when formants are sampled at 20% and 80% of the vowel duration versus only at steady-state, with no further improvement from sampling at additional points (Hillenbrand *et al.*, 1995). In addition, studies of silent-center vowels (e.g., Strange, 1989) have shown that listeners are able to identify vowels with reasonable accuracy even when given only the endpoints of the nucleus. The present results are consistent with the notion that the endpoints of the vowel trajectory are part of the phonetic signature of the vowel.

## 3. The clear speech effect for vowels

Three acoustic cues are generally agreed to drive the identity of vowels in English: F1 and F2 frequencies at steady-state, dynamic movement of these formants, and duration. Research examining the acoustic changes that underlie the superior intelligibility of vowels in clear speech has thus focused on these three acoustic characteristics, assuming that changes on these dimensions would explain the changes in intelligibility. In this study we have compared GLMMs with and without these acoustic measures, and found a significant although modest relative reduction of prediction error by 0.052. However, the results of the GLMM also revealed a significant clear speech benefit even after these measures had been accounted for, and the size of the remaining benefit was the same for the two listener groups. This indicates that other acoustic changes not captured by our measures contribute to vowel intelligibility, and

that they do so equally for YNH and EHI listeners. Possible candidates for these other changes include voice quality, fundamental frequency (either overall or the trajectory), formant bandwidth, or other aspects of the spectral envelope. Of these characteristics, only voice fundamental frequency has been compared in clear versus conversational speech, and only for sentence materials. Future research exploring these other changes will improve our understanding of vowel perception in general as well as in the details of what makes vowels more intelligible in clear speech.

Finally, it is important to note that although EHI listeners weighted duration and steady-state information differently from YNH listeners when identifying vowels, Ferguson (2012) reported and the GLMM confirmed that the size of the clear speech vowel intelligibility benefit was the same for the two groups. That is, EHI listeners seem to have adjusted their perceptual cue weighting so that they are able to take advantage of the acoustic differences between clear and conversational speech despite declines in signal audibility and changes in auditory and cognitive processing. Future research will use the Ferguson Database to explore how YNH and EHI listeners weight various acoustic cues when performing other perceptual tasks for clear and conversational speech, including identification of monosyllabic words and subjective ratings of sentence clarity.

## ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health Grant Nos. DC008886, DC005803, and HD002528. Katherine Beam, Jessica Stamey, Kyung Ae Keum, Billy Speer, Brett Adams, Craig Berg, Amanda Malone, Jill Freitas, and Patrick Pead assisted with acoustic analysis and data entry. Rajka Smiljanic provided very helpful comments on this manuscript. Development of the Ferguson Clear Speech Database was supported by National Institutes of Health Grant No. DC02229 to Indiana University.

<sup>1</sup>Although some authors have objected to using the term conversational to refer to read laboratory speech, we retain it for the sake of consistency with previous papers.

<sup>2</sup>Hazan and Baker (2011) recorded read conversational and clear speech in addition to eliciting semi-spontaneous conversational and clear speech using a novel technique called Diapix; all comments contained herein about that paper will refer to the read speech conditions.

ANSI (2012). ANSI S3.6-2012, Specifications for audiometers (American National Standards Institute, New York).

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics* (Cambridge University Press, Cambridge), pp. 241–302.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). “Mixed-effects modeling with crossed random effects for subjects and items,” *J. Mem. Lang.* **59**, 390–412.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). “Random effects structure for confirmatory hypothesis testing: Keep it maximal,” *J. Mem. Lang.* **68**, 255–278.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6, available at <http://CRAN.R-project.org/package=lme4> (Last viewed February 20, 2014).

Bradlow, A. R., Kraus, N., and Hayes, E. (2003). “Speaking clearly for children with learning disabilities: Sentence perception in noise,” *J. Speech Lang. Hear. Res.* **46**, 80–97.

- Coughlin, M., Kewley-Port, D., and Humes, L. E. (1998). "The relation between identification and discrimination of vowels in young and elderly listeners," *J. Acoust. Soc. Am.* **104**, 3597–3607.
- Ferguson, S. H. (2002). "Vowels in clear and conversational speech: Talker differences in acoustic features and intelligibility for normal-hearing listeners," Ph.D. dissertation, Indiana University.
- Ferguson, S. H. (2004). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," *J. Acoust. Soc. Am.* **116**, 2365–2373.
- Ferguson, S. H. (2012). "Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss," *J. Speech Lang. Hear. Res.* **55**, 779–790.
- Ferguson, S. H., and Kewley-Port, D. (2002). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **112**, 259–271.
- Ferguson, S. H., and Kewley-Port, D. (2007). "Talker differences in clear and conversational speech: Acoustic characteristics of vowels," *J. Speech Lang. Hear. Res.* **50**, 1241–1255.
- Ferguson, S. H., and Lee, J. (2006). "Vowel intelligibility in clear and conversational speech for cochlear implant users: A preliminary study," *J. Acad. Reh. Audiol.* **XXXIX**, 1–16.
- Ferguson, S. H., Poore, M. A., and Shrivastav, R. (2010). "Acoustic correlates of reported clear speech strategies," *J. Acad. Reh. Audiol.* **XLIII**, 45–64.
- Fox, R. A., and Jacewicz, E. (2009). "Cross-dialectal variation in formant dynamics of American English vowels," *J. Acoust. Soc. Am.* **126**, 2603–2618.
- Gagné, J. P., Masterson, V. M., Munhall, K. G., Bilida, N., and Querengesser, C. (1994). "Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech," *J. Acad. Reh. Audiol.* **27**, 135–158.
- Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, Cambridge), pp. 235–434.
- Gordon-Salant, S. (1986). "Recognition of natural and time-altered CVs by young and elderly subjects with normal hearing," *J. Acoust. Soc. Am.* **80**, 1599–1607.
- Gordon-Salant, S. (1987). "Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects," *J. Acoust. Soc. Am.* **81**, 1199–1202.
- Hazan, V., and Baker, R. (2011). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," *J. Acoust. Soc. Am.* **130**, 2139–2152.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications*, 2nd ed. (Lawrence Erlbaum, Mahwah, NJ), 392 pp.
- Janse, E. (2009). "Processing of fast speech by elderly listeners," *J. Acoust. Soc. Am.* **125**, 2361–2373.
- Krause, J. C., and Braid, L. D. (2004). "Acoustic properties of naturally produced clear speech at normal speaking rates," *J. Acoust. Soc. Am.* **115**, 362–378.
- Krause, J. C., and Braid, L. D. (2009). "Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech," *J. Acoust. Soc. Am.* **125**, 3346–3357.
- Kumar, U. A., and Sangamanatha, A. V. (2011). "Temporal processing abilities across different age groups," *J. Am. Acad. Audiol.* **22**, 5–12.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-3, available at <http://CRAN.R-project.org/package=lmerTest> (Last viewed February 24, 2014).
- Lam, J., Tjaden, K., and Wilding, G. (2012). "Acoustics of clear speech: Effect of instruction," *J. Speech. Lang. Hear. Res.* **55**, 1807–1821.
- Liénard, J. S., and Di Benedetto, M. G. (1999). "Effect of vocal effort on spectral properties of vowels," *J. Acoust. Soc. Am.* **106**, 411–422.
- Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F. G. (2004). "Clear speech perception in acoustic and electric hearing," *J. Acoust. Soc. Am.* **116**, 2374–2383.
- Liu, S., and Zeng, F. G. (2006). "Temporal properties in clear speech perception," *J. Acoust. Soc. Am.* **120**, 424–432.
- Molis, M. R., and Leek, M. R. (2011). "Vowel identification by listeners with hearing impairment in response to variation in formant frequencies," *J. Speech Lang. Hear. Res.* **54**, 1211–1223.
- Moon, S. J., and Lindblom, B. (1994). "Interaction between duration, context, and speaking style in English stressed vowels," *J. Acoust. Soc. Am.* **96**, 40–55.
- Neel, A. T. (2004). "Formant detail needed for vowel identification," *Acoust. Res. Lett. Online* **5**, 125–131.
- Payton, K. L., Uchanski, R. M., and Braid, L. D. (1994). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **95**, 1581–1592.
- Picheny, M. A., Durlach, N. I., and Braid, L. D. (1986). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," *J. Speech Hear. Res.* **29**, 434–446.
- Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-Effects Models in S and S-Plus* (Springer, New York), pp. 82–96.
- Quené, H. (2008). "Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo," *J. Acoust. Soc. Am.* **123**, 1104–1113.
- Quené, H., and Van den Bergh, H. (2004). "On multi-level modeling of data from repeated measures designs: A tutorial," *Speech Commun.* **43**, 103–121.
- Quené, H., and Van den Bergh, H. (2008). "Examples of mixed-effects modeling with crossed random effects and with binomial data," *J. Mem. Lang.* **59**, 413–425.
- R Core Team (2013). "R: A language and environment for statistical computing," Version 3.0.2. R Foundation for Statistical Computing, Vienna, Austria, available at <http://www.R-project.org> (Last viewed February 20, 2014).
- Rosen, K. M., Folker, J. E., Murdoch, B. E., Vogel, A. P., Cahill, L. M., Delatycki, M. B., and Corben, L. A. (2011). "Measures of spectral change and their application to habitual, slow, and clear speaking modes," *Int. J. Speech-Lang. Pathol.* **13**, 165–173.
- Schum, D. J. (1996). "Intelligibility of clear and conversational speech of young and elderly talkers," *J. Am. Acad. Audiol.* **7**, 212–218.
- Sjölander, K., and Beskow, J. (2006). "WaveSurfer," Stockholm: Centre for Speech Technology (CTT) at KTH, available at <http://sourceforge.net/projects/wavesurfer/> (Last viewed October 13, 2009).
- Smiljanic, R., and Bradlow, A. R. (2005). "Production and perception of clear speech in Croatian and English," *J. Acoust. Soc. Am.* **118**, 1677–1688.
- Smiljanic, R., and Bradlow, A. R. (2008). "Temporal organization of English clear and conversational speech," *J. Acoust. Soc. Am.* **124**, 3171–3182.
- Snijders, T. A. B., and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Sage Publications, London), 266 pp.
- Strange, W. (1989). "Dynamic specification of coarticulated vowels spoken in sentence context," *J. Acoust. Soc. Am.* **85**, 2135–2153.
- Trautman, H. (1990). "Analytical expressions for the tonotopic sensory scale," *J. Acoust. Soc. Am.* **88**, 97–100.
- Uchanski, R. M., Choi, S. S., Braid, L. D., and Durlach, N. I. (1996). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," *J. Speech Hear. Res.* **39**, 494–509.