Speech produced in noise: Relationship between listening difficulty and acoustic and durational parameters[a)]

Simone Graetzer,[1,b)] Pasquale Bottalico,[2] and Eric J. Hunter[2]
[1]Acoustics Research Unit, School of Architecture, University of Liverpool, Liverpool, England
[2]Voice Biomechanics and Acoustics Laboratory, Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan 48824, USA

(Received 10 February 2016; revised 17 May 2017; accepted 20 July 2017; published online xx xx xxxx)

Conversational speech produced in noise can be characterised by increases in intelligibility relative to such speech produced in quiet. Listening difficulty (LD) is a metric that can be used to evaluate speech transmission performance more sensitively than intelligibility scores in situations in which performance is likely to be high. The objectives of the present study were to evaluate the LD of speech produced in different noise and style conditions, to evaluate the spectral and durational speech modifications associated with these conditions, and to determine whether any of the spectral and durational parameters predicted LD. Nineteen subjects were instructed to speak at normal and loud volumes in the presence of background noise at 40.5 dB(A) and babble noise at 61 dB(A). The speech signals were amplitude-normalised, combined with pink noise to obtain a signal-to-noise ratio of −6 dB, and presented to twenty raters who judged their LD. Vowel duration, fundamental frequency and the proportion of the spectral energy in high vs low frequencies increased with the noise level within both styles. LD was lowest when the speech was produced in the presence of high level noise and at a loud volume, indicating improved intelligibility. Spectrum balance was observed to predict LD.
[http://dx.doi.org/10.1121/1.4997906]

[MAH] Pages: 1–10

I. INTRODUCTION

Talkers modify their speech in the presence of noise to maintain a level that is sufficient for communication. The Lombard effect (Lombard, 1911) is the involuntary tendency to increase the level of speech in the presence of noise. In noisy environments, speakers commonly increase not only their vocal intensity but also their fundamental frequency (f_o; e.g., Junqua, 1993; Van Summers *et al.*, 1988), their first vowel formant ($F1$; e.g., Boril and Pollak, 2005; Kadiri, 1998), and the energy in the spectrum between 1 and 4 kHz relative to the energy below 1 kHz, resulting in an increase in the *spectrum balance* (hereafter SB; e.g., Stanton *et al.*, 1988; Ternström *et al.*, 2006; Krause and Braida, 2004, 2009; Lu and Cooke, 2009). As the speech level and, therefore, the vocal effort level increases, the spectrum flattens (e.g., Nordenberg and Sundberg, 2004; Ternström *et al.*, 2006). When this level increase co-occurs with an increase in glottal flow and, hence, subglottal pressure, f_o typically rises, while $F1$ rises with more jaw opening (Fant, 1997). Speech produced in noise can also demonstrate changes in segment (especially vowel) duration and/or a slowing of the speech rate (e.g., Fonagy and Fonagy, 1966; Junqua, 1993; Krause and Braida, 2004).

So-called Lombard speech is more intelligible than speech produced in quiet environments when presented at equal signal-to-noise (SN) ratios (e.g., Dreher and O'Neill, 1957; Summers *et al.*, 1988; Pittman and Wiley, 2001; Lu and Cooke, 2009). However, it has not yet been resolved which of the speech modifications contribute most strongly or are necessary, either in combination or in isolation, to a gain in intelligibility (whether linguistic or non-linguistic parameters; cf. Cooke and García Lecumberri, 2012). See Cooke *et al.* (2014a) for a review. Relatedly, it has not yet been determined which of these speech modifications predict perceived listening difficulty (LD; e.g., Morimoto *et al.*, 2004). However, an upward shift in the overall spectral "centre of gravity" (CoG) does appear to contribute more to intelligibility than does an f_o increase (e.g., Hazan and Markham, 2004; Lu and Cooke, 2009; Mayo *et al.*, 2012) or the sorts of durational changes that occur in Lombard speech (Cooke *et al.*, 2014b).

Under some conditions, noise-induced speech modifications may be harmful to intelligibility in quiet conditions. Findings for both shouted speech (e.g., Pickett, 1956; Junqua, 1993) and non-native listeners (Cooke and García Lecumberri, 2012) indicate that the speech level may be increased to preserve audibility to the detriment of phonetic information (Rostolland and Parant, 1975). Junqua (1993) observed variation in the intelligibility of Lombard speech relative to speech produced in quiet, depending on the vocabulary, noise type (white Gaussian or multi-talker) and talker gender. For non-native vs native listeners, Cooke and García Lecumberri (2012) found that Lombard speech may be slightly less intelligible than conversational speech when presented in quiet. However, Lombard speech may provide benefits to both native and non-native listeners by placing important speech information outside of the range

---

[a)] An earlier version of this study was presented at the 45th Annual Symposium of the Voice Foundation, Philadelphia, PA, June 2016.
[b)] Electronic mail: s.graetzer@liverpool.ac.uk

of the energetic masker (see discussion in García Lecumberri *et al.*, 2010; Cooke *et al.*, 2014a; Godoy *et al.*, 2014; ISO 226, 2003).

Previous research has shown that there are differences in speech intelligibility between Lombard and "clear" speech or interlocutor-directed speech, such as speech directed toward infants, hearing-impaired persons, and non-native speakers (e.g., Picheny *et al.*, 1985, 1986; Skowronski and Harris, 2006; Wassink *et al.*, 2007; Godoy *et al.*, 2014; Cooke *et al.*, 2014a). Some clear speech modifications are enhancements that are dependent on linguistic knowledge, and therefore favour the native speaker (e.g., Picheny *et al.*, 1986; Bond and Moore, 1994; Bradlow and Bent, 2002; Hazan and Markham, 2004).

The acoustic and durational differences between Lombard and loud or shouted speech have been considered by Stanton and colleagues (e.g., Stanton, 1988; Stanton *et al.*, 1988), and Bond and Moore (1990), but much remains to be investigated. Stanton (1988) compared the speech modifications associated with the change from normal speech to speech produced at "nominally 10 dB above normal," to the modifications associated with the change from normal to Lombard speech (involving 90 dB of pink noise being emitted into the talker's ears via headphones) in the fighter cockpit environment. He noted a smaller shift in the spectral CoG and $F1$ and, typically, a smaller increase in vowel duration between normal and Lombard speech than between normal and loud speech, although there was large inter-speaker variation. Bond and Moore (1990) concluded, based on a single speaker's production, that Lombard speech and deliberately loud speech (involving an instruction concerning imagined speaker-listener distance) result from the same speech production mechanisms.

Intelligibility assessment in speech communication can be performed by means of objective and subjective methods, such as by calculating the Speech Transmission Index (STI) of a transmission channel (IEC 60268-16, 2011), or by testing with real listeners the percentage of words correctly understood within a given space (intelligibility scores, or IS). However, sentence scores of 100% are associated in the ISO 9921 (2003) standard with a large range of STI values (0.45–1). A sentence intelligibility score of 100% does not imply that each word is clearly understood, and there are many situations in which the speech transmission performance cannot be regarded as satisfactory (ISO 9921, 2003; Morimoto *et al.*, 2004, p. 1609). Additionally, for the same communication channel, scores can be high while predictability and/or word familiarity are high, but reduce when words are unpredictable or unfamiliar (e.g., Kalikow *et al.*, 1977). These issues of metric sensitivity in the context of high performance and highly familiar and/or predictable speech material can be resolved with the use of a rating scale concerning how difficult a given listening situation is (e.g., ITU-T P.85, 1994; IEC 60268-16, 2011). LD is a subjective perception metric developed by Morimoto, Sato and colleagues for use with highly familiar words that can be used to evaluate speech transmission performance more accurately and sensitively than IS in situations in which the performance is likely to be high (Morimoto *et al.*, 2004). It is designed to minimise the potential confounding effects of word familiarity and predictability and the extent of higher cognitive processing. LD ratings using the 0–3 rating system described by Sato and colleagues (Sato *et al.*, 2005) are mapped to IS and STI values in IEC 60268-16E (2011). LD has been used as a complement to IS or the STI in several publications concerning the transmission of Japanese or Korean speech (e.g., Morimoto *et al.*, 2004; Sato *et al.*, 2005; Lee and Jeon, 2011).

While the listening effort scale has been expanded from 5 to up to 13 or more levels in order to avoid floor or ceiling saturation effects (ITU-T P.85, 1994), the LD traditionally has only 4 levels (from not difficult to extremely difficult), and is defined as the percentage of the total number of responses that indicates some level of difficulty. The use of only four levels can lead to an accumulation of values at the upper bound (Morimoto *et al.*, 2004; Genta *et al.*, 2013), while averaging over the total number of responses means that variability associated with the individual listener's responses cannot be modeled. This variability, which can be high (see, e.g., Lee and Jeon, 2011; Genta *et al.*, 2013), may be due to individual differences in cognitive ability or preference. Studies of category scale design have indicated that data quality (e.g., reliability, sensitivity) tends to improve as the number of answer categories increases (e.g., Alwin, 1992). An alternative seven-point scale for rating LD was proposed by Gover and Bradley (2007), and a five-point scale attempting to address the saturation issue but not the variation issue was proposed by Genta *et al.* (2013), who suggested on the basis of their results that there was a need for alternative implementations of the method. A ten-point scale LD metric and statistical approach designed to address both issues of saturation and listener variation is presented in this paper. An additional contribution of the paper is the use of LD ratings with first language English speakers and listeners who have been audiometrically tested for normal hearing.

The consideration of LD independently from speech intelligibility is particularly important for hearing aid users, young children, and older listeners. This is because even under conditions of perfect speech intelligibility, adverse conditions such as background noise can impair memory of spoken items and listening comprehension (e.g., Pichora-Fuller, 2003). The literature indicates that there are many acoustical modifications of speech associated with ease of listening that may or may not co-occur with improved intelligibility such as modifications of speech rate f_o, formant frequencies, and f_o modulation (Bond and Moore, 1994; Lu and Cooke, 2009; Cooke *et al.*, 2014a). Decreased LD is likely to reduce listener fatigue, which may lead to intelligibility improvements in extended listening tasks (Lim and Oppenheim, 1979). In contexts in which listening is difficult, acoustic treatments or signal enhancement may be used to reduce fatigue and improve recall (Lim and Oppenheim, 1979).

In summary, there has been much work contributing to the understanding of speech in noise, and features of clear speech and interlocutor-directed speech. However, the question of which speech modifications inherently improve intelligibility and reduce LD for the normal hearing native English speaker has not yet been fully resolved. Moreover, there is a need for further investigation and modification of the LD metric.

In the present study, the principal aim was to evaluate acoustic and durational modifications that occur in non-communicative laboratory speech in noisy environments in two different speech styles and to relate them to perceived LD. Based on this aim, the primary research question was as follows, where speech was produced in babble noise at 61 dB(A) or in background noise at 40.5 dB(A): Do any of the spectral or durational measures considered—f_o (in semitones), f_o modulation, SB, or vowel duration—predict LD ratings when SN-ratio = −6 dB? A further aim was to extend previous results (Morimoto *et al.*, 2004; Sato *et al.*, 2005) that indicated that LD may be a useful measure of rating speech transmission when word recognition performance would be high. This new work extends the previous work by using first-language normal hearing English talkers and listeners and also by including an evaluation of which spectral or durational changes predict LD ratings for continuous speech produced in noise. In this way, the current paper responds to a call for studies examining the relationship between the intonational, spectral, and durational features of Lombard speech and speech intelligibility (Cooke *et al.*, 2014b). Furthermore, while Lu and Cooke (2009) considered the relevant contributions of only f_o and spectrum flattening parameters, and suggested that durational increases might, like spectrum flattening, contribute to intelligibility, in the current study, f_o, spectrum flattening, and a durational parameter are considered. It is worth considering whether the acoustical parameters that predict speech intelligibility also predict listening difficulty. The findings have implications for the improvement of communication in noisy environments.

II. EXPERIMENTAL PROCEDURES

This study was conducted with approval from and in accordance with the policies of Michigan State University's Human Research Protection Program (IRB No. 13-1149). Participants were not compensated. MATLAB v2014b and Praat v5.4.01 (Boersma and Weenink, 2015) were used for signal processing. Post-processing and statistical analysis were conducted in R v3.1.2 (R Development Core Team, 2016).

A. Experiment one: Speech assessment

1. Subjects and instructions

Nineteen native American English speaking subjects (nine males, ten females) of between 18 and 29 years of age with a mean of 21 years of age and with self-reported normal speech and hearing were recruited. The subjects were recorded while reading the "Rainbow" passage text in a semi-reverberant room (a classroom) in two different styles, corresponding to normal and loud voice levels. In the environment of the talker was multi-talker children's babble (classroom babble; *high level*) noise and/or (naturally occurring) background (*low level*) noise, which was primarily associated with the heating, ventilation, and air conditioning system. The instructions given for the styles were as follows: normal: "Speak in your normal voice" and loud: "Imagine you are in a classroom and you want to be heard by all of the children." Investigators were present in the room, observing the talker.

2. Room acoustic measurements and pre-processing

The recording took place in a classroom of dimensions 5.8 m × 6 m × 2.7 m. The floor and ceiling were covered by absorbent material (carpet and absorbent tiles). Room acoustic parameters were measured in an unoccupied state without furniture from the impulse responses (IRs) generated by balloon pops (according to ISO 3382-2, 2008). T_{30} was derived by means of the AURORA software suite (Farina, 2010). The mid-frequency reverberation time was 0.53 s (standard deviation = 0.04; see Bottalico *et al.*, 2015).

The background noise was measured in the unoccupied room using a Head and Torso Simulator (HATS) Kemar 45BB-1 (G.R.A.S., Denmark). The primary noise source contributing to the level of 40.5 dB(A) in the talker position was the ventilation system. Given that the level was below 43 dB(A), the level of speech production in the background noise condition was not affected by the noise (Lazarus, 1986; Bottalico *et al.*, 2017).

The multi-talker noise was emitted by a directional loud speaker (Yamaha studio monitor model HS5, Yamaha, Japan) at a level of 61 dB(A) in the talker position. This level represents a common noise level (hereafter, Lnoise) generated by children in a classroom engaged in quiet group work or individual work with some movement (Shield and Dockrell, 2004). The spectral maxima in the babble occurred in the 500 Hz and 1 kHz octave bands. Babble noise was emitted by the loud speaker rather than by headphones to avoid the perturbation of the talkers' self-monitoring of auditory feedback. Arguably, if noise is delivered via headphones, the headphones can alter the talker's perception of their own voice (due to the effects on both internal and external hearing), and therefore the talker's voice production (e.g., Garnier and Henrich, 2014). The babble signal had deep amplitude fluctuations, while the background noise was stationary. The mean f_o of the babble was 256 Hz, which is within the normal range for children (Titze, 2000). In the babble noise condition, the SN-ratio of the speech signal (represented by the concatenated voiced segments) and corresponding noise signal as acquired by the head-mounted microphone was estimated at +24 dB on average in the loud style, and +22 dB on average in the normal style.

The speech signal was acquired by an omnidirectional head-mounted microphone (Glottal Enterprises M-80, Syracuse, NY) placed at a distance of 5 cm from the mouth (much less than the critical distance; hence, the signal was associated only with the direct sound of the talker). The microphone has a fairly flat frequency response <4 kHz, with a rising frequency response between 4 and 6 kHz, and a sensitivity of −65 dB ± 3 dB. The signal was acquired by a Roland R-05 digital recorder (Hamamatsu, Japan) in 16 bit/44.1 kHz WAV format. The microphone line out was connected to a personal computer (PC) via an external sound board (Scarlett 2i4 Focusrite, High Wycombe, UK). The signal was recorded with Audacity v2.0.6 with a sampling rate of 44.1 kHz. Recordings varied in length between 25 and 45 s, depending on the talker.

Words were manually segmented in Praat. For the vowel duration analysis, individual vowels were segmented in Python v3.4 by means of the FAVE-align and HTK toolkits and visually inspected for errors. The FAVE-align toolkit is an adaptation of the Penn Forced Aligner, which relies on hidden Markov modeling (Rosenfelder *et al.*, 2014). Vowels were labeled according to the Carnegie Mellon University (CMU) Pronouncing Dictionary representations of the relevant Rainbow passage words.

3. Vowel duration

Normalised vowel duration was calculated by dividing each vowel duration in seconds associated with a given subject by that subject's mean in the low Lnoise and normal style (a presumed baseline value). Due to heteroscedasticity, durations were analysed by means of Welch-corrected one-way tests for equal means. Speech rate was considered during testing, but was found not to change in a reliable way with the level of noise and so is excluded from the analyses.

4. Fundamental frequency

f_o was extracted from the recordings by means of Praat at 10 ms intervals. An autocorrelation-based method was used with Hanning windows with a length of 0.043 s, a pitch floor of 70 Hz, and a pitch ceiling of 400 Hz. f_o was then converted to semitones in R with bases for males and females equal to their mean f_o (Hz): 128 Hz for males and 203 Hz for females in this case. These base values are representative of typical adult males and females, the difference relating primarily to differences in membranous vocal fold length (Titze, 2000, 2011).

5. Spectrum balance

Sound pressure level (SPL) data concerning the same talkers and experimental conditions as in the present study have been reported in a previous publication (Bottalico *et al.*, 2015). In this previous study, concerning a set of speech production data of which the present data are a subset, it was confirmed that SPL increased in speech produced in noisy conditions, specifically, unintelligible children's babble at 61 dB(A), relative to speech produced in relatively quiet conditions [ambient noise at 40.5 dB(A)]. As in Bottalico *et al.* (2015), in the present study, MATLAB version 2014b was used to obtain a time history of overall SPL evaluated at 0.125 s intervals for each reading of the Rainbow passage. The average among all the SPL values was computed per subject and this mean was subtracted from each time history value for that subject (termed ΔSPL). This within-subject centering was performed in order to evaluate the variation in the subject's vocal behaviour in the different conditions from their typical vocal behaviour. For each subject, the relative amplitudes in each octave band were calculated in dB, where each data point corresponded to a difference between each level measured in dB for a subject and the maximum amplitude calculated for that subject across noise and style conditions.

Spectral analysis was conducted in order to determine whether an increase in the SB occurred in high relative to low Lnoise. SB, named after the measure of Ternström *et al.* (2006; but modified in form), was defined as the energy difference between the 1–4 kHz and 0–1 kHz regions or *bands* (i.e., the mean energy computed for the upper band minus the mean computed for the lower band, in dB). The upper band limits were chosen on the basis of previous studies (e.g., Krause and Braida, 2004, 2009; Garnier and Henrich, 2014). The SB value will usually be negative, as the low frequency region tends to dominate the voice spectrum. The SB increases when it goes from more to less negative and, thus, becomes less steep (or in other words, the spectrum becomes more flat). The claim is that in intelligible speech produced by normal talkers, the energy difference between the lower and the upper bands becomes smaller, resulting in an increase in the SB. However, as discussed previously, this difference can also be affected by the speech level and f_o. SB as here defined relates to the α ratio measure (but with the negative rather than the positive sign and an upper limit of 4 kHz rather than 5 kHz; see, e.g., Sundberg and Nordenberg, 2006).

In order to measure possible measurement bias due to any babble noise in the signal acquired by the head-mounted microphone, the difference in SB with and without the artificial babble noise for the same speech material was evaluated with a HATS. The same speech material recorded in the same experimental conditions was emitted from the mouth simulator, with and without babble noise being emitted by the loud speaker. The average difference in the SB with and without the babble noise was equal to 0.12 ± 1.14 dB. A paired sample *t*-test indicated that this difference was negligible [$t = -0.67$, degrees of freedom (df) $= 49$, $p = 0.51$].

The concatenated words (i.e., the sentences with silences between words removed) produced by each talker in each condition were subjected to long term average spectrum (LTAS) analysis, also performed in Praat. After fast Fourier analysis, each LTAS was calculated and the SB was derived via the "get slope" function with the lower band limits of 0 and 1 kHz, and the upper band limits of 1 and 4 kHz, where the energy is averaged over the concatenated signal in dB, based on the mean power of the signal. When the results were compared with those produced with a lower band of 50 Hz–1 kHz, the difference was negligible.

An evaluation of the effects of noise, style, and interactions of noise and style, noise and gender, and style and gender on the response variable, SB, was conducted by means of a linear mixed effects or LME model (*lme4* and *lmerTest* R packages) fitted by restricted maximum likelihood (REML) with the random effects term of talker. The LME model output includes the estimates of the fixed effects coefficients, the standard error (SE) associated with the estimate, the df, the test statistic, *t*, and the *p* value. The Satterthwaite method is used to approximate df and calculate *p* values.

B. Experiment two: LD assessment

Prior to the LD assessment, 20 native American English speaking listeners (10 males, 10 females), who were aged between 18 and 23 years, with a mean age of 21 years, were audiometrically assessed to ensure normal hearing at ≤20 dB hearing level (HL) between 250 Hz and 6 kHz using an Orbiter 922 v. 2 audiometer (Madsen Kft., Budapest, Hungary) audiometer in a sound-attenuated booth.

1. Room acoustic measurements and pre-processing

There were 152 test stimuli per listener (19 talkers × 2 speaking styles × 2 noise conditions × 2 external auditory feedback conditions, which are not considered here). The stimuli were prepared as follows. A short extract of the Rainbow passage (two sentences in length, which did not include the first or the last phrases in the passage) produced by each talker in each condition was linearly amplitude normalised and combined with pink noise in MATLAB to obtain a SN-ratio of -6 dB. This value is the lowest considered by Sato *et al.* (2005). The onset of noise preceded the onset of the signal by 500 ms. The background Lnoise in the listener position in the booth was 25.1 dB(A), as measured using an NTi Measurements microphone M2211 (class 1 frequency response) and analysed by means of NTi XL2 Audio and Acoustic Analyzer (Schaan, Liechtenstein). LD ratings have been used previously with a specific short speech pattern (Kurisu *et al.*, 2013).

2. Testing procedures

The stimuli were presented binaurally via Sennheiser HD205 headphones (Wedemark, Germany) in a pseudo-random order to 20 listeners seated in a sound-attenuated booth. Randomisation on the order of presentation and the recording of LD ratings was obtained via a custom Praat script. The instruction was "rate the level of LD for these sentences on a scale of 1 (not difficult, no effort required) to 10 (very difficult, considerable effort required)." Testing was divided into a training phase (8 stimuli) and a testing phase (152 stimuli), and subjects were able to rest between the 2 halves of the testing phase, to reduce any effects of fatigue. The training phase was included and exposure of all listeners to all conditions was specified, in part, to minimise possible context effects (see Sato *et al.*, 2005). The LD assessment took $\sim$45 min. Subjects were required to respond to every stimulus.

In the current study, the discrete subjective LD scale was changed from 1 to 4 (as in the original 2004 version of the metric), in which the percentage of values >1 are taken to represent the LD associated with a given experimental condition (Morimoto *et al.*, 2004) to 1 to 10, for reasons outlined in Sec. II B 1.

3. Statistical procedures

A cumulative link mixed model (Laplace approximation; *ordinal* R package) was run with LD as the response variable and Lnoise, style, their interaction, and interactions of both Lnoise and style with talker gender, with both the listener and the talker as random effects terms. To determine which, if any, of the acoustic and durational parameters predicted LD, a LME model fitted by REML was run with LD as the response variable and SB, f_o (semitones), f_o (semitones) standard deviation, and normalised vowel duration as independent variables, with an interaction of f_o (semitones) and gender, and with talker as the random effects term. In the case of this model, LD was averaged across listeners per signal. Given that the resolution of 1/10 and the SN-ratio of -6 dB led to the LD metric having good coverage of the measurand range, this response variable could be treated as continuous.

III. RESULTS

First, the effects of Lnoise and style on spectral and durational speech parameters will be reported. Second, the extent to which any of these parameters predict LD will be discussed.

A. Experiment one: Speech assessment

1. Vowel duration

Welch-corrected one way tests for equal means indicated that there was an effect of Lnoise [$F(1{,}18649) = 134.44$, $p < 0.0001$], and gender [$F(1{,}18985) = 15.75$, $p < 0.0001$] but not style ($p > 0.1$) on normalised vowel duration. This effect of Lnoise held per style and per vowel quality [/i/, $F(1{,}1062) = 7.25$, $p < 0.01$; /a/, $F(1{,}149) = 4.83$, $p < 0.05$; /u/, $F(1{,}528) = 6.60$, $p < 0.05$]. As shown in Fig. 1, vowel durations were longer when the speech was produced in high level than low Lnoise, for both males and females.

2. Fundamental frequency

The mean f_o increased from 200 to 207 Hz from low to high Lnoise for females, and from 125 to 131 Hz from low to high Lnoise for males. Not only the males' but also the females' mean f_o remained distant from the mean f_o of the babble signal (256 Hz).

A LME model was built with f_o (semitones) as the response variable, and as predictors: Lnoise, style, and interactions of Lnoise and style and noise and gender. Talker was included as a random factor. The low Lnoise, the normal style, and the male gender were chosen as the reference levels. As is shown in Fig. 2, f_o (in semitones) was higher when speech was

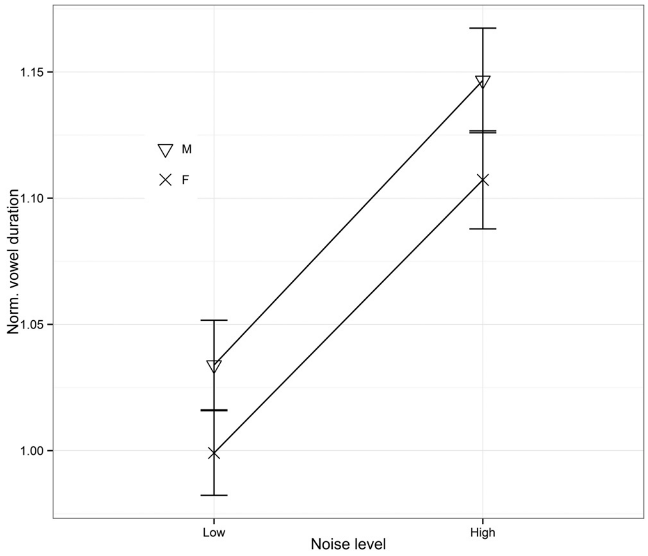

FIG. 1. Normalised vowel durations by Lnoise (x axis) and gender (symbol) condition. Means are shown with 95% confidence intervals.

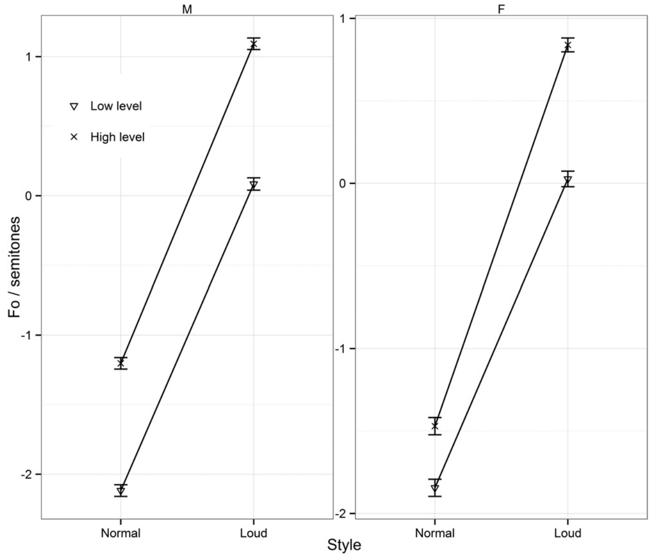

FIG. 2. F_o in semitones per style (x axis), Lnoise (symbol) and gender [(left) male, (right) female] condition. Means are shown with 95% confidence intervals.

produced in the presence of high Lnoise than low Lnoise ($\hat{\beta} = 0.76$, SE = 0.03, df = 254566, $t = 29.70$, $p < 0.0001$). F_o was higher in the loud style than in the normal style ($\hat{\beta} = 2.09$, SE = 0.03, df = 254566, $t = 81.83$, $p < 0.0001$). There was an interaction between noise and style ($\hat{\beta} = 0.33$, SE = 0.03, df = 25466, $t = 11.52$, $p < 0.0001$) such that the effect of noise was stronger in the loud style. There was also an interaction between style and gender ($\hat{\beta} = -0.11$, SE = 0.03, df = 25466, $t = -3.85$, $p < 0.001$), such that males increased their f_o more than females in the loud relative to the normal style.

In the normal style, variation in f_o (semitones) in the form of standard deviations was slightly increased when speech was produced in the presence of high Lnoise than low Lnoise ($\hat{\beta} = 0.33$, SE = 0.14, df = 129, $t = 2.32$, $p < 0.05$). In the loud style, f_o variation did not appear to be reliably associated with noise conditions. Variation tended to be lower in the loud style than in the normal style ($\hat{\beta} = -0.22$, SE = 0.11, df = 129, $t = -1.96$, $p = 0.05$). Very similar results were found when the f_o values were subjected to outlier detection and removal using the Bonferroni method before analysis, indicating that these results were not due to f_o artefacts.

3. SB

With regard to within-subject normalised overall SPL (ΔSPL), in the normal style, ΔSPL increased by approximately 9 dB from −11.74 dB in low Lnoise to −2.70 dB in high Lnoise. In the loud style, ΔSPL increased by approximately 4.70 dB from 5.03 dB in low Lnoise to 9.71 dB in high Lnoise. The relative magnitude of spectral energy in the higher frequencies was increased in the high Lnoise, as indicated by the relative amplitudes (dB) in each of the seven octave bands (Fig. 3). For males, there tended to be a smaller difference between Lnoise conditions in the loud style than in the normal style, as in the case of the overall ΔSPL.

Amplitude variation, measured in terms of the range of the relative amplitude, increased from the low to the high Lnoise in the normal style by 3 dB, and in the loud style for the females by 2 dB, but did not increase in the loud style for males, possibly due to a ceiling effect.

The effects of noise, style, and gender on SB are reported in Table I and shown in Fig. 4. Recall that SB will typically

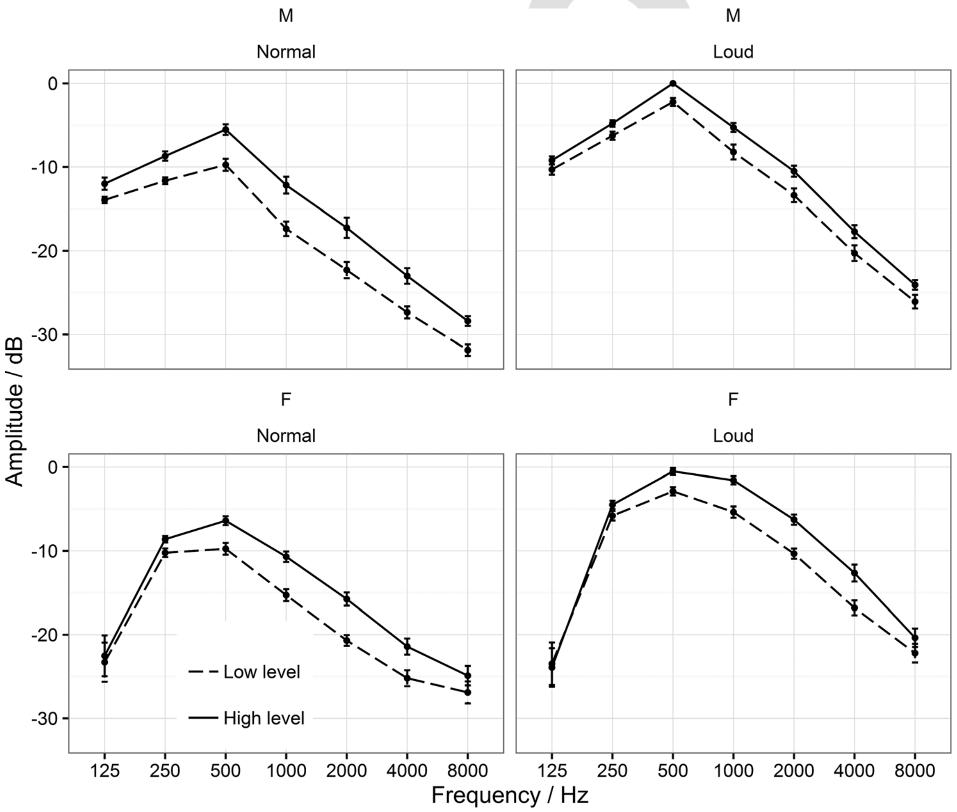

FIG. 3. Relative amplitude (dB) by frequency (Hz), style [(left) loud, (right) normal] and noise level (dashed line, low; solid line, high) for males (upper) and females (lower) with ± 1 SE.

TABLE I. LME model with the response variable SB and independent variables Lnoise, and style and interaction terms with gender (reference levels: Lnoise, low; style, normal; gender male). Significance codes: *** < 0.001, ** < 0.01, * < 0.05, "." < 0.1.

Term	Estimate	SE	df	*t*
(Intercept)	−17.60	0.55	21	−31.26***
Lnoise high	1.75	0.27	129	6.46***
Style loud	3.31	0.27	129	8.88***
Lnoise high: Style loud	−0.72	0.31	129	−2.34*
Lnoise low: Gender female	0.76	0.75	20	1.02
Lnoise high: Gender female	1.30	0.75	20	1.74
Style loud: Gender female	1.73	0.31	128	5.62***

TABLE II. Cumulative link mixed model (Laplace) output for LD by Lnoise and style and interactions with talker gender (reference levels are Lnoise, low; style, Normal; gender male). Significance codes: *** < 0.001, ** < 0.01, * < 0.05, "." < 0.1.

Term	Estimate	SE	*z*
Lnoise high	−0.86	0.12	−7.40***
Style loud	−1.17	0.12	−10.02***
Lnoise high: Style loud	0.57	0.13	4.37***
Lnoise low: Gender female	−0.61	0.35	−1.74.
Lnoise high: Gender female	−0.61	0.35	−1.73.
Style loud: Gender female	−0.16	0.13	−1.20

increase (become less negative) as SPL increases. The LME model included Lnoise and style and interactions of Lnoise and style, Lnoise and gender, and style and gender, with talker as a random effect. Reference levels were low Lnoise, normal style, and male gender. When the speech was produced in high vs low Lnoise in the normal and in the loud styles, there was an increase in SB. In addition, when the speech was produced in the loud style vs the normal style in the presence of low Lnoise, there was an increase in SB. In the normal style, there was a greater difference between Lnoise conditions than in the loud style. This result suggests the achievement of an upper limit in the high Lnoise, loud style condition. There was also an interaction between style and gender: For males there was a much smaller difference between the style conditions than for the females (Fig. 4). For females, f_o was moderately positively correlated with SB ($r = 0.52$, $p < 0.0001$). No reliable correlation was observed for males ($r = 0.24$, $p < 0.05$).

B. Experiment two: LD assessment

1. Effects of noise, style, and talker gender on LD

In the speech perception study, 20 listeners evaluated their difficulty in listening to the speech produced by the talkers in the 2 noise and 2 style conditions. A cumulative link mixed model was fit to LD with the following predictors: Lnoise and style and interactions of Lnoise and style, Lnoise and gender, and style and gender. The model incorporated random effects for talker and listener and for the talker listener interaction. The reference levels were low Lnoise, normal style, and male gender. As reported in Table II and shown in Fig. 5, there was a decrease in LD when the speech was produced in high vs low Lnoise in the normal style, and when the speech was produced in the loud style vs normal style in low Lnoise. There was an interaction of noise and style such that the difference in LD between the styles was greater in low Lnoise than in high Lnoise. The lowest LD scores occurred when speech was produced in high Lnoise in the loud style condition.

When LD was converted by quartile to a four-point scale (as in the original method), the effects of noise and style were very similar to those observed in the ten-point scale model. In the four-point scale model, the arcsine transformed proportion of values higher than one (averaged over the listeners) was evaluated as the response variable of a LME model with noise and style and their interaction as independent variables, and talker as a random factor.

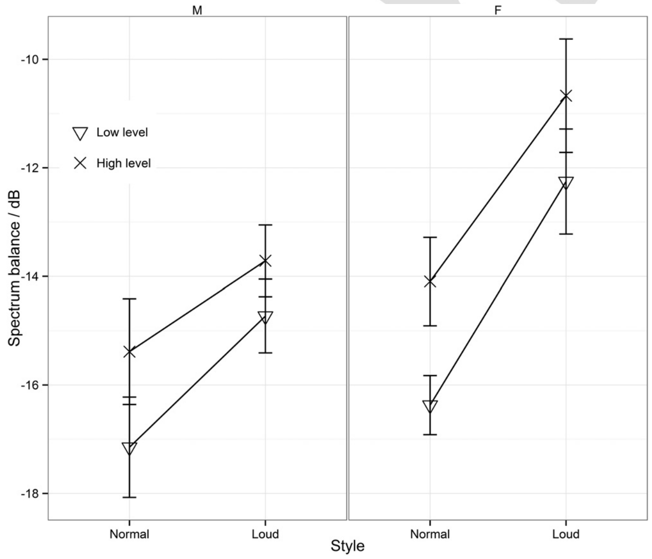

FIG. 4. SB in dB per style (x axis), noise (symbol), and gender [(left) male, (right) female] condition, with means and 95% confidence intervals.

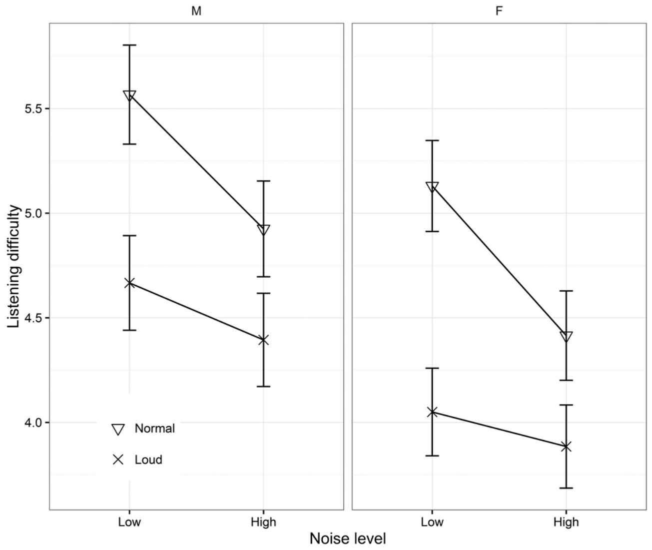

FIG. 5. LD (1, lowest; 10, highest) by Lnoise (x axis) and style (symbol) condition, with means and 95% confidence intervals.

TABLE III. LME model with the response variable LD (averaged over talker) and scaled independent variables: SB, f_o modulation (semitones), vowel duration, and an interaction of f_o and talker gender. Significance codes: *** < 0.001, ** < 0.01, * < 0.05, "." < 0.1.

Term	Estimate	SE	df	*t*
(Intercept)	0.05	0.77	135	0.07
SB	−2.9	0.03	82	−10.87***
F_o standard deviation (semitones)	0.11	0.08	50	1.46
Vowel duration	−0.14	0.52	146	0.27
F_o (semitones): Gender male	−0.01	0.04	56	−0.27
F_o (semitones): Gender female	0.13	0.06	71	2.38*

2. Relationships between speech parameters and LD

Models were fitted to determine which of the acoustic and durational parameters predicted LD. The distribution of the LD ratings was near normal, with no saturation at the upper and lower bounds. The results are reported in Table III. A LME model was run with LD as the response variable and the acoustical and durational parameters as independent variables: SB, f_o modulation (semitones), normalised vowel duration, and an interaction of f_o and talker gender. Of the parameters, only SB reliably predicted LD ($p < 0.0001$). However, there was a difference in the slope f_o (semitones)-LD between females and males such that for females there was a decrease in LD as f_o increased ($p < 0.05$). The slope can be derived from a simple linear regression:

$$y = 4.27–0.15 f_o + \epsilon. \quad (1)$$

IV. DISCUSSION

This paper reports the use of LD ratings for an identification of the speech modifications that predict the transmission performance of speech produced in noise by first-language, normal hearing English speakers. In the assessment of the speech parameters in this study, the increase in vocal intensity in speech produced in noise was found to co-occur with increases in f_o and SB, as predicted on the basis of previous studies (e.g., Van Summers *et al.*, 1988; Stanton *et al.*, 1988; Junqua, 1993). Arguably, these spectral modifications, which occurred in a non-communicative context, are primarily associated with the increase in vocal intensity in the presence of babble noise, but could also reflect other modifications made to improve audibility for the talker at his/her own ear (see, e.g., Garnier and Henrich, 2014; Cooke *et al.*, 2014a).

In the present study, it was possible to identify effects of noise within both speech styles. First, for normalised vowel duration, there was an effect of Lnoise but no observable effect of style. Additionally, for f_o, SB, and LD, there was an additive effect of Lnoise and style. In full, the effects of noise in the environment of the talker were an increase in vowel duration, an increase in f_o, an increase in the SB, and, in the perception assessment, a decrease in ratings of LD.

The results concerning the relationship between durational changes and LD ratings suggest that while vowel elongation can increase the amount of acoustic information available about vowel quality and neighbouring segment identity (see, e.g., Fonagy and Fonagy, 1966), the extent to which these changes can improve the intelligibility of speech masked by broadband noise appears to depend on other factors (Cooke *et al.*, 2014b; Lu and Cooke, 2009). The magnitude of the vowel duration results may reflect the fact that the high Lnoise present during speech production was multi-talker noise, which is said to degrade the perception of vowels more than consonants (Junqua, 1993). It is interesting that there was no reliable effect of style on vowel duration for these speakers, despite the observed increase in speech level from normal to loud style (cf., e.g., Traunmüller and Eriksson, 2000).

Shifts in the spectral energy distribution toward frequencies between 1 and 4 kHz, i.e., increases in SB, were observed to predict LD when the signals were presented to listeners at the same SN-ratio. The reported effects of these shifts on LD ratings are consistent with the results of Krause and Braida (2004), who found that high frequency spectral emphasis contributes to the increased intelligibility of clear relative to conversational speech when produced in noise. In the present study, it was found that while changes in both f_o and spectral energy distribution occur when speech is produced in noise, only the latter appears to contribute in a significant way to intelligibility (Lu and Cooke, 2009; Hazan and Markham, 2004; Cooke *et al.*, 2014b). The f_o increase may under most conditions merely accompany the increase in vocal intensity (Gramming *et al.*, 1988). Lu and Cooke (2009) have argued that SB, unlike an upward shift in f_o, reliably increases the amount of information available to the listener, i.e., the amount of speech information that is out of the range of the masker energy. In Cooke's (2006) glimpsing model of speech perception in noise, there are more glimpses (defined as connected regions in the spectro-temporal representation of the speech time-frequency plane) within which speech information is audible. In other words, in the current study, an increase in SB provides some release from *energetic* masking. In singers, an increase in energy in the region of 3 kHz allows the voice to be heard well above an orchestra or background noise (Sundberg, 1994) and results in an increase in phons and sones (Hunter *et al.*, 2006), due to the human ear being particularly sensitive to frequencies in this region (see Cooke *et al.*, 2014a; ISO 226, 2003). As mentioned previously, Cooke and García Lecumberri (2012) have argued that while some linguistic enhancements may exist in Lombard speech, such as greater vowel space dispersion (Cooke and Lu, 2010), these may be outweighed by other changes that in fact reduce intelligibility; linguistic enhancements, therefore, appear to have a limited role.

With regard to f_o, in the normal style, a small but reliable increase was observed in f_o modulation (in semitones) in high vs low Lnoise, which has previously been interpreted as evidence of an active strategy to improve audibility in noise (Garnier and Henrich, 2014). For females, the increase in f_o (semitones) with the increase in Lnoise was larger in the loud than in the normal style, despite the effect of noise on speech level being larger in the normal style, which may also suggest an active strategy to optimise masker release. Further, for the female speakers, a possible explanation of the finding for the slope f_o (semitones)-LD may be that within the f_o range of the females ($\sim$160–270 Hz), when f_o increases there may be some additional release from energetic masking. This is due to the migration of spectral energy into higher parts of the spectrum (given the wider spacing of harmonics at high f_o frequencies). This release may be associated with the presence of a high level of noise and/or the raised intensity of the loud style (see, e.g., Cooke *et al.*, 2014a). Such a relationship between f_o and speech intelligibility for female talkers may only occur at low SN-ratios (Barker and Cooke, 2007). Indeed, in the current study, for females but not males, f_o was moderately positively correlated with SB.

The increase in amplitude variation in high Lnoise relative to low Lnoise for some speakers may not be entirely related to the increase in vocal intensity, but may also reflect the intention of these speakers to improve their intelligibility once the SN-ratio can no longer be improved (e.g., Picheny *et al.*, 1985, 1986; Ternström *et al.*, 2006).

On the basis of the results presented, it can be argued that LD ratings are sensitive to changes in the audibility or intelligibility of speech in contexts in which the performance would be high due, in part, to the predictability of the speech material rather than strictly to the SN-ratio or the characteristics of the masker. The results are consistent with the findings of Morimoto *et al.* (2004) that "LD is not always high when background noise is present." (p. 1611) This research suggests that an artificial increase in the SB, for example, generated by a filter that amplifies frequencies > 1 kHz, may reduce LD. Such processing is feasible to implement in real-time (Skowronski and Harris, 2006). Thus, signals may be enhanced to improve comprehension and recall for young children and older listeners. In this paper, a revised LD method has been presented that addresses the issues of saturation at ceiling performance and high listener variability that have been reported in the literature.

V. CONCLUSIONS

The objectives of the present study were to evaluate the LD of speech produced in different noise and style conditions, evaluate the spectral and durational speech modifications associated with these conditions, and determine whether any of the spectral and durational parameters predicted LD. It was confirmed that speech produced in high level babble noise relative to low level background noise was associated with an increased f_o, increased spectral energy between 1 and 4 kHz relative to energy below 1 kHz, and increased vowel duration. However, only the proportion of high to low spectral energy reliably predicted LD for normal-hearing listeners.

It should be noted that the speech was acquired in the high Lnoise condition in the presence of babble noise; however, the effects of noise in the acquired signal on SB itself, being the difference between the mean energy of the 1–4 kHz band and of the <1 kHz band, were negligible. In this study, the ecological validity of tests in terms of proprioception and internal and external auditory feedback was prioritised, as was unconstrained head movement in the loud style (see, e.g., Lagier *et al.*, 2010; Garnier and Henrich, 2014).

Further studies are required to evaluate the ten-point scale form of the LD measure. In a future study, the properties of the original form, the revised form, and IS will be compared both for repeated and unique speech material. Moreover, not only the level but also the type of noise in the talker's environment will be manipulated during communicative tasks, for example, among broadband, speech-shaped, and babble noise, to allow a clear separation of the effects on speech audibility and intelligibility of the level from the type of noise. The type of additive noise used in the listening experiment will also be varied to evaluate how LD ratings and word recognition scores are affected by the properties of the noise masker.

ACKNOWLEDGMENTS

Thanks are due to the subjects who participated in the study and to members of the audience who were present when an earlier version of this work was presented at the Voice Symposium in June 2016. We also wish to thank the associate editor and two anonymous reviewers for their constructive feedback. This research was primarily supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) of the National Institutes of Health (NIH) under Grant No. R01DC012315. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Alwin, D. F. (**1992**). "Information transmission in the survey interview: Number of response categories and the reliability of attitude measurement," Sociol. Methodol. **22**, 83–118.

Barker, J., and Cooke, M. (**2007**). "Modelling speaker intelligibility in noise," Speech Commun. **49**, 402–417.

Boersma, P., and Weenink, D. (**2015**). "PRAAT: Doing phonetics by computer (version 5.4.01) [computer program]," http://www.praat.org (Last viewed 1 May 2015).

Bond, Z. S., and Moore, T. J. (**1990**). "A note on Loud and Lombard speech," in *Proc. 1st International Conference on Spoken Language Processing (ICSLP)*, Kobe, Japan, pp. 969–972.

Bond, Z. S., and Moore, T. J. (**1994**). "A note on the acoustic-phonetic characteristics of inadvertently clear speech," Speech Commun. **14**, 325–337.

Boril, H., and Pollak, P. (**2005**). "Design and collection of Czech Lombard speech database," in *Proc. Interspeech 2005*, Lisboa, Portugal, pp. 1577–1580.

Bottalico, P., Graetzer, S., and Hunter, E. J. (**2015**). "Effects of voice style, noise level, and acoustic feedback on objective and subjective evaluations," J. Acoust. Soc. Am. **138**(6), 498–503.

Bottalico, P., Ipsaro Passione, I., Graetzer, S., and Hunter, E. J. (**2017**). "Evaluation of the starting point of the Lombard effect," Acta Acust. Acust. **103**(1), 169–172.

Bradlow, A., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Cooke, M., and García Lecumberri, M. L. (**2012**). "The intelligibility of Lombard speech for non-native listeners," J. Acoust. Soc. Am. **132**(2), 1120–1129.

Cooke, M., King, S., Garnier, M., and Aubanel, V. (**2014a**). "The listening talker: A review of human and algorithmic context-induced modifications of speech," Comput. Speech Lang. **28**, 543–571.

Cooke, M., and Lu, Y. (**2010**). "Spectral and temporal changes to speech produced in the presence of energetic and informational maskers," J. Acoust. Soc. Am. **128**(4), 2059–2069.

Cooke, M., Mayo, C., and Villegas, J. (**2014b**). "The contribution of durational and spectral changes to the Lombard speech intelligibility benefit," J. Acoust. Soc. Am. **135**(2), 874–883.

Dreher, J. J., and O'Neill, J. J. (**1957**). "Effects of ambient noise on speaker intelligibility for words and phrases," J. Acoust. Soc. Am. **29**(12), 1320–1323.

Fant, G. (**1997**). "The voice source in connected speech," Speech Commun. **22**, 125–139.

Farina, A. (**2010**). "Aurora plug-ins," available at http://www.aurora-plugins.com (Last viewed 20 January 2016).

Fonagy, I., and Fonagy, J. (**1966**). "Sound pressure level and duration," Phonetica **15**, 14–21.

García Lecumberri, M. L., Cooke, M., and Cutler, A. (**2010**). "Non-native speech perception in adverse conditions: A review," Speech Commun. **52**, 864–886.

Garnier, M., and Henrich, N. (**2014**). "Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise," Comput. Speech Lang. **28**, 580–597.

Genta, G., Astolfi, A., Bottalico, P., Barbato, G., and Levi, R. (**2013**). "Management of truncated data in speech transmission evaluation for pupils in classrooms," Measur. Sci. Rev. **13**(2), 75–82.

Godoy, E., Koutsogiannaki, M., and Stylianou, Y. (**2014**). "Approaching speech intelligibility enhancement with inspiration from Lombard and clear speaking styles," Comput. Speech Lang. **28**(2), 629–647.

Gover, B. N., and Bradley, J. S. (**2007**). "Comparison of subjective and objective ratings of intelligibility of speech recordings," Can. Acoust. **35**(3), 140–141.

Gramming, P., Sundberg, J., Ternstrom, S., Leanderson, R., and Perkins, W. (**1988**). "Relationship between changes in voice pitch and loudness," J. Voice **2**, 118–126.

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**(5), 3108–3118.

Hunter, E. J., Švec, J. G., and Titze, I. R. (**2006**). "Comparison of the produced and perceived voice range profiles in untrained and trained classical singers," J. Voice **20**(4), 513–526.

IEC (**2011**). 60268-16(E). "Sound system equipment, Part 16: Objective rating of speech intelligibility by speech transmission index" (European Committee for Standardization, Brussels, Belgium, 2011).

ISO 226:2003(E) (**2003**). "Acoustics—Normal equal-loudness level contours" (International Organization for Standardization, Geneva, Switzerland).

ISO 9921:2003E (**2003**). "Ergonomics: Assessment of speech communication" (International Organization for Standardization, Geneva, Switzerland).

ISO 3382-2:2008(E) (**2008**). "Acoustics—Measurement of room acoustic parameters, Part 2: Reverberation time in ordinary rooms" (International Organization for Standardization, Geneva, Switzerland).

ITU-T P.85 (**1994**). "A method for subjective performance assessment of the quality of speech voice output devices," ITU-T Recommendation P.85 (International Telecommunication Union, Geneva, Switzerland).

Junqua, J. C. (**1993**). "The Lombard reflex and its role on human listeners and automatic speech recognizers," J. Acoust. Soc. Am. **93**, 510–524.

Kadiri, N. (**1998**). "Conséquences d'un environnement bruité sur la production de la parole" ("Consequences of a noisy environment for speech production"), Ph.D. dissertation, Toulouse University, France.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**(5), 1337–1351.

Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**(1), 362–378.

Krause, J. C., and Braida, L. D. (**2009**). "Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech," J. Acoust. Soc. Am. **125**(5), 3346–3357.

Kurisu, K., Nakamura, S., Yasu, K., and Arai, T. (**2013**). "Signal to overlap-masking ratio of the broadcasted speech and its listening difficulty: An application for an evaluation tool of sound system tuning," Acoust. Sci. Tech. **34**(5), 354–355.

Lagier, A., Vaugoyeau, M., Ghio, A., Legou, T., Giovanni, A., and Assaiante, C. (**2010**). "Coordination between posture and phonation in vocal effort behaviour," Folia Phoniatr. Logop. **62**, 195–202.

Lazarus, H. (**1986**). "Prediction of verbal communication in noise—A review: Part 1," Appl. Acoust. **19**, 439–463.

Lee, P. J., and Jeon, J. Y. (**2011**). "Evaluation of speech transmission in open public spaces affected by combined noises," J. Acoust. Soc. Am. **130**(1), 219–227.

Lim, J. S., and Oppenheim, A. V. (**1979**). "Enhancement and bandwidth compression of noisy speech," Proc. IEEE **67**(12), 1586–1604.

Lombard, É. (**1911**). "Le signe de l'élévation de la voix" ("The sign of the elevation of the voice"), Ann. Maladies de L'Oreille Larynx **37**(2), 101–119.

Lu, Y., and Cooke, M. (**2009**). "The contribution of changes in *F*0 and spectral tilt to increased intelligibility of speech produced in noise," Speech Commun. **51**, 1253–1262.

Mayo, C., Aubanel, V., and Cooke, M. (**2012**). "Effect of prosodic changes on speech intelligibility," in *Proc. 13th Annual Conference of the International Speech Communication Association (Interspeech)*.

Morimoto, M., Sato, H., and Kobayashi, M. (**2004**). "Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces," J. Acoust. Soc. Am. **116**(3), 1607–1613.

Nordenberg, M., and Sundberg, J. (**2004**). "Effect on LTAS of vocal loudness variation," in *Speech, Music and Hearing I, Quarterly Progress Status Report 45* (Royal Institute of Technology, Stockholm), pp. 93–100.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," J. Speech. Lang. Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," J. Speech. Lang. Hear. Res. **29**, 434–446.

Pichora-Fuller, M. K. (**2003**). "Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension," Int. J. Audiol. **42**, S59–S67.

Pickett, J. (**1956**). "Effects of vocal force on the intelligibility of speech sounds," J. Acoust. Soc. Am. **28**(5), 902–905.

Pittman, A. L., and Wiley, T. L. (**2001**). "Recognition of speech produced in noise," J. Speech. Lang. Hear. Res. **44**, 487–496.

R Development Core Team (**2016**). "R: A language and environment for statistical computing," available at http://www.R-project.org (Last viewed 1 May 2016).

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., and Yuan, J. (**2014**). FAVE (Forced Alignment and Vowel Extraction) Program Suite Version 1.1.3 10.5281/zenodo.9846, available at http://www.research.ed.ac.uk/portal/en/publications/fave-forced-alignment-and-vowel-extraction-suite-version-113(bbc2046d-6768-47c5-b574-2987895b0307).html.

Rostolland, D., and Parant, C. (**1975**). "The intelligibility of a foreign language in a noisy environment," in *Proc. Symposium Intelligibility of Speech*, Nottingham, UK.

Sato, S., Bradley, J. S., and Morimoto, M. (**2005**). "Using listening difficulty ratings of conditions for speech communication in rooms," J. Acoust. Soc. Am. **117**(3), 1157–1167.

Shield, B., and Dockrell, J. E. (**2004**). "External and internal noise surveys of London primary schools," J. Acoust. Soc. Am. **115**(2), 730–738.

Skowronski, M. D., and Harris, J. G. (**2006**). "Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments," Speech Commun. **48**, 549–558.

Stanton, B. J. (**1988**). "Robust recognition of loud and Lombard speech in the fighter cockpit environment," Ph.D. dissertation, Purdue University, West Lafayette, IN.

Stanton, B. J., Jamieson, L. H., and Allen, G. D. (**1988**). "Acoustic-phonetic analysis of loud and Lombard speech in simulated cockpit conditions," in *Proc. 13th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol. 1, pp. 331–334.

Sundberg, J. (**1994**). "Perceptual aspects of singing," J. Voice **8**(2), 106–122.

Sundberg, J., and Nordenberg, M. (**2006**). "Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech," J. Acoust. Soc. Am. **120**(1), 453–457.

Ternström, S., Bohman, M., and Södersten, M. (**2006**). "Loud speech over noise: Some spectral attributes, with gender differences," J. Acoust. Soc. Am. **119**(3), 1648–1665.

Titze, I. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech, Iowa), pp. 1–409.

Titze, I. (**2011**). "Vocal fold mass is not a useful quantity for describing *F*0 in vocalization," J. Speech. Lang. Hear. Res. **54**(2), 520–522.

Traunmüller, H., and Eriksson, A. (**2000**). "Acoustic effects of variation in vocal effort by men, women, and children," J. Acoust. Soc. Am. **107**(6), 3438–3451.

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effect of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**(3), 917–928.

Wassink, A. B., Wright, R. A., and Franklin, A. D. (**2007**). "Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers," J. Phon. **35**(3), 363–379.