The intelligibility of Lombard speech for non-native listeners

Martin Cooke[a)]
Language and Speech Laboratory, Universidad del País Vasco, 01006 Vitoria, Spain

Maria Luisa García Lecumberri
Departamento de Filologia Inglesa, Facultad de Letras, Universidad del Pais Vasco, 01006 Vitoria, Spain

(Received 19 October 2011; revised 6 June 2012; accepted 13 June 2012)

Speech produced in the presence of noise—Lombard speech—is more intelligible in noise than speech produced in quiet, but the origin of this advantage is poorly understood. Some of the benefit appears to arise from auditory factors such as energetic masking release, but a role for linguistic enhancements similar to those exhibited in clear speech is possible. The current study examined the effect of Lombard speech in noise and in quiet for Spanish learners of English. Non-native listeners showed a substantial benefit of Lombard speech in noise, although not quite as large as that displayed by native listeners tested on the same task in an earlier study [Lu and Cooke (2008), J. Acoust. Soc. Am. **124**, 3261–3275]. The difference between the two groups is unlikely to be due to energetic masking. However, Lombard speech was less intelligible in quiet for non-native listeners than normal speech. The relatively small difference in Lombard benefit in noise for native and non-native listeners, along with the absence of Lombard benefit in quiet, suggests that any contribution of linguistic enhancements in the Lombard benefit for natives is small.
 [http://dx.doi.org/10.1121/1.4732062]

PACS number(s): 43.71.Hw [CGC] Pages: 1120–1129

I. INTRODUCTION

When faced with noise, talkers modify the way they speak. Over a century ago, Lombard (1911) reported that a patient presented with noise immediately increased his vocal effort and fundamental frequency. In the intervening years many studies have confirmed these basic findings for English (e.g., Junqua, 1993; Summers *et al.*, 1988; Hansen, 1996) and for other languages such as French (Garnier, 2007) and Spanish (Castellanos *et al.*, 1996), and extended them to include increases in first formant frequency, an upwards shift of spectral center of gravity and overall segment lengthening. Critically, "Lombard" speech has been found to be more intelligible than speech produced in quiet conditions when tested in additive noise (Dreher and O'Neill, 1957; Summers *et al.*, 1988; Pittman and Wiley, 2001; Garnier, 2007; Lu and Cooke, 2008), sometimes by substantial amounts. For instance, Lu and Cooke (2008) observed an increase in keyword scores of 25 percentage points, corresponding to a relative gain of 59%, for Lombard over normal speech when presented in speech-shaped noise at a signal-to-noise ratio (SNR) of −9 dB. Given the increasing use of speech output technology—whether synthetic, recorded or delayed live speech—in everyday conditions where noise is present, it is of interest to discover the origins of the Lombard speech intelligibility benefit in order to inform the development of speech modification strategies which promote robust communication.

To investigate possible causes of the Lombard intelligibility benefit, Lu and Cooke (2009) modified speech that had been produced in quiet by mapping two parameters—fundamental frequency (F0) and spectral tilt—to values observed in Lombard speech. Spectral tilt changes were responsible for about two-thirds of the intelligibility benefit, while F0 modification produced no gains when applied alone, nor additional gains when used in combination with spectral tilt changes. Using the glimpsing model (Cooke, 2006), Lu and Cooke (2009) suggest that the Lombard intelligibility benefit is derived in large part from the reduction in energetic masking that occurs in frequency regions of importance for speech perception due to an upwards shift in the overall spectral center of gravity which is typically observed in Lombard speech (Fig. 2 in Sec. III illustrates the long-term spectral profiles of normal and Lombard speech).

Lombard speech can be seen as a response to noise, but speech production also changes as a result of explicit instruction (Chen, 1980; Picheny *et al.*, 1985; Cutler and Butterfield, 1990; Payton *et al.*, 1994). When asked to speak clearly, talkers typically respond with global modifications (e.g., slower speech rate, higher F0, and larger F0 range) as well as both acoustic-phonetic (e.g., vowel space expansion, increases in consonant-vowel energy ratio) and phonological changes (e.g., fewer vowel reductions and fewer instances of alveolar tapping in English). The resulting "clear speech" has been shown to be beneficial in noise to normal hearers as well as listeners with hearing impairment and non-native listeners, although—as for Lombard speech—the contribution of each of the observed modifications is not yet fully understood (Uchanski, 2005; Smiljanic and Bradlow, 2009).

While even the briefest acquaintance with Lombard and clear forms of speech reveals that they are quite different, and that Lombard speech in particular gives the informal impression of being out of place when heard in quiet conditions, the possibility exists that Lombard speech shares some

[a)]Author to whom correspondence should be addressed. Also at: Ikerbasque (Basque Science Foundation). Electronic mail: m.cooke@ikerbasque.org

 0001-4966/2012/132(2)/1120/10/$30.00

of the "linguistic enhancements"—the acoustic-phonetic and phonological changes highlighted above—exhibited by clear speech, and that these contribute to the Lombard intelligibility advantage. The potential for linguistic enhancements certainly exists in Lombard speech, since spectral, durational and other modifications are not constant across the speech signal but differ at the level of speech segments (e.g., Junqua, 1993; Lu and Cooke, 2008). Further, it is known that vowel space modifications are present in Lombard speech (Bond *et al.*, 1989; Garnier, 2007; Bořil, 2008; Cooke and Lu, 2010). Some of these vowel space changes are similar in Lombard and clear speech styles. For instance, a reduction in within-category dispersion has been observed for both Lombard (Cooke and Lu, 2010) and clear (Chen, 1980) speech. Differences also exist: vowel space expansion typically seen in clear speech and known to benefit listeners (Bradlow *et al.*, 1996) was not present for Lombard speech (Cooke and Lu, 2010). Nevertheless, the existence of acoustic-phonetic and phonological changes in Lombard speech motivates the hypothesis that some of the Lombard advantage is due to linguistic enhancements.

Since the intelligibility of Lombard speech is measured in the presence of masking noise, it can be difficult to distinguish the possible contributions to the Lombard advantage from energetic masking release and linguistic enhancements. Acoustic-phonetic or phonological changes may confer benefits by placing speech information out of the range of masker energy, either as a deliberate strategy on the part of the speaker, or as a side-effect of other modifications. For example, increases in vowel duration may lead to greater resistance to masking by affording more epochs where formant information is audible, especially for nonstationary maskers. In general, *any* modification to speech has the potential to affect its susceptibility to energetic masking.

The clear speech intelligibility benefit has been evaluated using non-native groups in a series of studies by Bradlow and her colleagues. Bradlow and Bent (2002) found that while native listeners displayed a clear speech benefit over conversational speech of 16 rationalized arcsine units (RAUs), a non-native group benefited by significantly less, around 5 RAUs, suggesting that at least some clear speech enhancements are language-specific and their exploitation depends on linguistic knowledge. More recently, Smiljanic and Bradlow (2011) reported that highly proficient non-native listeners found clear speech enhancements as beneficial as native listeners, which they consider to further support the idea that clear speech strategies are to some extent language-specific and thus require a good command of the linguistic code in order to exploit them effectively.

In a similar vein, the current study attempts to distinguish the roles of energetic masking release and linguistic enhancement through the use of non-native listeners. This listener group is of interest for two reasons. First, there is now strong evidence, reviewed in García Lecumberri *et al.* (2010), that energetic masking affects native and non-native listeners in equal amounts for speech material with limited semantic content, simple syntax and common words (see also Cutler *et al.*, 2004). Thus, we hypothesize that if native listeners possess a greater Lombard advantage in noise than non-native listeners, then this will be due to the contribution of factors other than energetic masking release. Second, since even proficient non-native listeners typically perform well below ceiling in noise-free conditions (Black and Hast, 1962; Cooke *et al.*, 2008), it is possible to make a direct estimate of any Lombard benefit which does not depend on energetic masking release, by comparing their scores for normal and Lombard speech in quiet.

There are very few reported studies of the effect of Lombard speech on non-native listeners. Junqua (1993) tested French, British English, and American English listener groups on isolated American English words (alphanumeric and control words) recorded in quiet and in white-Gaussian noise at 85 dB SPL. Intelligibility tests were carried out using these words mixed with white-Gaussian noise at different SNRs for each listener group (+10, 0, and −10 dB, respectively). All three listener groups showed no Lombard advantage in this study. Indeed, apart from the spoken nasals "en" and "em," Lombard speech was less intelligible than speech recorded in quiet conditions. The absence of a Lombard benefit in this case may be due to the type of speech material or masking noise employed. Since the three listener groups were tested at widely differing SNRs, it is difficult to draw any strong conclusions about the comparative intelligibility of Lombard speech in noise for native and non-native listeners. More recently, Li (2004) reported in a brief abstract on a study involving normal and Lombard sentences spoken by English and Cantonese speakers and recorded in 70 dB of cafeteria noise. Cantonese listeners transcribed sentences presented in quiet and at an unspecified SNR. In noise, Lombard speech from both Cantonese and English speakers showed higher transcription errors than normal speech. No results were reported for the quiet presentation condition. Again, it is not clear why no Lombard advantage was present in this study.

The origins of the Lombard intelligibility advantage are explored in the current study through two research questions. The first concerns the existence and size of the Lombard benefit in noise for non-native listeners and how it compares to the native benefit. Since we argue that energetic masking effects will be similar for the two groups, we hypothesize that any difference in Lombard benefit is an indicator of the presence and scale of other contributions to the Lombard benefit, including linguistic enhancements. The second question relates to the Lombard benefit enjoyed by non-native listeners in quiet. The degree of benefit or otherwise provides a direct measure of Lombard-normal differences that cannot be attributed to energetic masking release.

The two experiments of the current study used the same speech/masker materials and task employed in an earlier study with native listeners (Lu and Cooke, 2008). Apart from the availability of native listener scores on the same task, this corpus was chosen because it resulted in very large Lombard benefits in noise, unlike the two aforementioned studies which compared native and non-native perception of Lombard speech (Junqua, 1993; Li, 2004). Additionally, an earlier study using the same type of lexically and syntactically simple sentence material in stationary noise (Cooke *et al.*, 2008) demonstrated a constant native advantage in

quiet and masked conditions, confirming that the use of higher-level knowledge known to benefit native listeners (e.g., Gat and Keith, 1978; Meador *et al.*, 2000; García Lecumberri *et al.*, 2010) is minimized for these materials.

Experiment I measured the intelligibility of keywords in simple British English normal and Lombard sentences by Spanish learners of English. Listeners were tested in quiet and in the presence of masking noise. A second experiment replicated key conditions of experiment I, both to rule out speech subset effects and to quantify the influence of the masker spectrum on the size of the Lombard advantage.

II. EXPERIMENT I: INTELLIGIBILITY OF NORMAL AND LOMBARD SPEECH IN QUIET AND NOISE FOR NON-NATIVE LISTENERS

Experiment I evaluated non-native listeners' keyword identification rates in quiet and in the presence of a speech-shaped noise (SSN) masker at a range of SNRs. Sentences were simple utterances produced by eight talkers in quiet and in two levels of noise. Speech and noise materials were the same as those used to test Lombard benefits in native listeners (Lu and Cooke, 2008), allowing a direct comparison of the two listener groups.

A. Listeners

Fifty-seven listeners (48 female, 9 male, age: 19–39, mean: 20.9 years) all second year undergraduates studying English Philology at the University of the Basque Country, Spain, took part in experiment I. All students enrolled on the course had passed English grammar exams corresponding to a B2 (Upper Intermediate) level according to the Common European Framework of Reference for Languages (CEF) and were, at the time of the experiment, taking language courses at the next level (C1, Advanced). All received course credit for their participation.

Nine listeners were excluded from the analysis for the following reasons. One listener reported hearing problems, while two listeners had a first language other than Spanish or Basque.[1] A further five listeners produced responses which were outliers (defined here as falling outside ± 1.5 times the interquartile range) in more than one condition, while one listener produced a response which was a severe outlier in one condition (a score of 10% in a condition with an across-listener mean of 80%). The analysis reported here is based on the remaining 48 listeners.

B. Speech and noise materials

The speech material used in this study was a subset of utterances employed in Lu and Cooke (2008). In that study, sentences were drawn from the Grid Corpus (Cooke *et al.*, 2006), which defines simple six-word utterances such as "place red at G9 again" and "lay blue with X4 soon." Grid sentences allow for all combinations of three keywords denoting one of four colors, one of 25 spoken English letters (excluding the multisyllabic "W") and the 10 spoken digits. Here, as in Lu and Cooke (2008), the listeners' task was to report the alphanumeric pair of keywords (e.g. "G9"). Four male and four female talkers produced a random selection of Grid sentences in quiet and in the presence of noise delivered over headphones. In the current study, three subsets of data were used. One set of utterances was produced in quiet conditions, which we refer to here as "normal." In a further two conditions, speech was produced in the presence of speech-shaped noise derived from normal speech from the Grid Corpus (Cooke *et al.*, 2006) with presentation levels of 82 and 96 dB SPL, which we refer to as "Lombard82" and "Lombard96."[2] Each of the eight talkers produced 50 normal sentences and 50 sentences in each of the Lombard conditions.

Listeners were tested in quiet and with SSN added at SNRs of 0, −5, and −9 dB. The SSN sample was the one used in Lu and Cooke (2008) and had a long-term spectrum matching that of the normal (i.e., non-Lombard) speech in the Grid Corpus (see Fig. 2). The −9 dB value was chosen since that was used to measure native responses to normal and Lombard speech in Lu and Cooke (2008). However, due to the possibility that non-native listeners might reach close to floor performance at this noise level, two less intense noise levels were also used in an attempt to bracket the absolute performance levels shown by native listeners (viz: 42% for normal speech, 64% and 67% for the two levels of Lombard speech). Noise was co-gated with sentence stimuli and 10 ms half-Hamming ramps were applied to minimize onset and offset artefacts. Sentences mixed with noise as well as sentences presented in quiet were normalized to the same RMS energy on presentation. Each experimental block (one of three speech types and one of four noise levels, including quiet) contained 50 stimuli drawn at random from the appropriate speech type.

C. Procedure

The experiment took place in a quiet language laboratory. Listeners heard stimuli over Plantronics Audio-90 headphones, delivered via a custom MATLAB program. After each stimulus, participants responded by pressing a letter key followed by a number key (0–9) on their computer keyboard. The experiment was self-paced: After responding to the current stimulus, participants heard the next stimulus after a short pause. No correct-answer feedback was given. Listeners were familiarized with the task and keyboard in a short practice session containing 15 unscored exemplars each of normal and Lombard speech. The order of stimulus blocks was randomized across participants and the stimulus order within each condition was also randomized. Listeners adjusted the overall volume to a comfortable listening level during a practice phase, after which no further changes in volume were made.

D. Results

Listener responses were scored as percentages of letter and digit keywords correctly identified, with percentages converted to rationalized arcsine units (RAU; Studebaker, 1985) for all subsequent statistical analyses. However, for ease of interpretation, results are displayed as percentages and differences in percentage points.

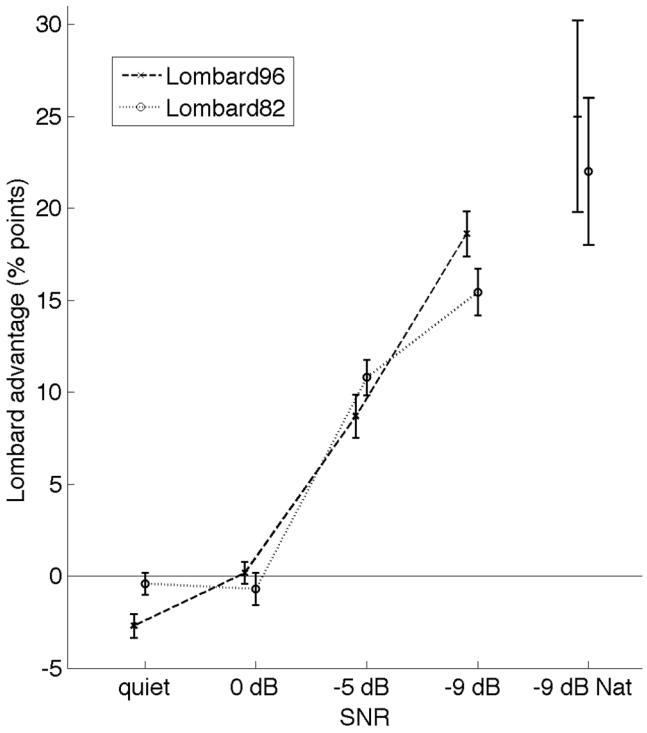

FIG. 1. The Lombard advantage (keyword score for Lombard speech minus keyword score for normal speech) in percentage points as a function of SNR and inducer noise level. The rightmost data points indicate native Lombard advantages in the −9 dB condition, taken from Lu and Cooke (2008). Absolute keyword scores for normal speech are 93.5% (quiet), 81.5% (0 dB), 62.0% (−5 dB), and 36.2% (−9 dB). Error bars here and elsewhere indicate ±1 standard errors.

Figure 1 shows the difference in the percentage of keywords correct between each Lombard condition and normal speech, which we refer to as the Lombard advantage. At high noise levels (−5 and −9 dB SNR) non-native listeners showed a clear intelligibility gain for both Lombard82 and Lombard96 over speech produced without noise. The Lombard advantage reaches 19 percentage points for Lombard speech induced by 96 dB noise when presented at −9 dB SNR, from a baseline of 36% for normal speech, while for Lombard82 the benefit is around 15 percentage points. For comparison, native listeners tested in Lu and Cooke (2008, Fig. 5) showed gains of 25 and 22 percentage points for Lombard96 and Lombard82, respectively, at −9 dB SNR, from a baseline for normal speech of 42%.

However, the Lombard advantage is clearly seen to be dependent on noise level: the benefit in the −5 dB SNR condition is reduced relative to the −9 dB level, and disappears altogether at 0 dB SNR. In the quiet condition, Lombard96 speech is in fact *less* intelligible than normal speech $[t(47) = 4.83, p < 0.001]$, while Lombard82 has equivalent intelligibility to normal speech in quiet $[p = 0.67]$. A repeated-measures ANOVA with within-subjects factors of presentation noise level (quiet, 0 dB, −5 dB, −9 dB) and speech type (normal, Lombard82, Lombard96) confirmed the presentation level x speech type interaction $[F(6, 282) = 48.5, p < 0.001, \eta^2 = 0.19]$ as well as main effects of level $[F(3, 141)] = 1945, p < 0.001, \eta^2 = 0.85]$ and speech type $[F(2, 94) = 92.3, p < 0.001, \eta^2 = 0.11]$.

While at a SNR of −9 dB Lombard96 speech is more intelligible than Lombard82, Fig. 1 suggests a tendency at −5 dB SNR for Lombard speech induced by the less intense noise to be more beneficial than that induced by more intense noise $[t(47) = 1.97, p = 0.056]$. It may be that Lombard speech resulting from moderate noise is better matched to produce intelligibility gains in more moderate noise levels, although further studies using a wide range of SNRs and Lombard speech noise presentation levels are needed to explore this possibility.

E. Interim discussion

Experiment I extends to non-native listeners the finding that, in noise, Lombard speech is substantially more intelligible than speech produced in quiet conditions. This outcome suggests that factors which promote Lombard intelligibility in noise are also of value to participants listening in a second language. The Lombard advantage was greater at more adverse noise levels, as found with native listeners by Summers *et al.* (1988). One possible explanation is that the probability of masking of important information-bearing elements of speech, such as those conveyed by the location of the second formant, increases with noise level, and since Lombard speech shifts energy to the mid-frequencies, speech information in these regions is better able to escape masking than normal speech at more adverse SNRs.

For normal speech, the difference between native and non-native listener scores was about 6 percentage points at a SNR of −9 dB. Since non-native listeners obtained scores of 93.5% in quiet where native performance is close to ceiling, the absolute native advantage is seen to be similar in quiet and noisy conditions, supporting previous findings reviewed in García Lecumberri *et al.* (2010) and suggesting that energetic masking affected both listener groups by equivalent amounts in the current task. By comparison, for Lombard speech in noise the native advantage was 12 percentage points. Therefore, it seems likely that the additional native benefit of 6 percentage points derives from factors other than energetic masking release, including possible linguistic enhancements. This outcome parallels the finding that native listeners derive a greater clear speech benefit than non-native listeners.

However, a further key finding from experiment I is the absence of a non-native Lombard benefit in quiet for non-native listeners. Indeed, the results here suggest that Lombard speech induced by intense noise is *less* intelligible than normal speech in the absence of a masker. Thus, it may be that non-native listeners derive no benefit from putative linguistic enhancements in Lombard speech, calling into question the scale of such enhancements and their contribution for native listeners too. Since the difference between native and non-native Lombard benefits in noise is relatively small, and given that non-natives actually suffer from Lombard speech in quiet, it seems reasonable to speculate that native listeners would derive relatively little benefit from Lombard speech in quiet. Of course, native listeners can be expected to be performing at or very near ceiling in that condition. The Appendix reports on a test of normal and

Lombard speech in quiet with a separate cohort of native listeners which found a greater number of errors for Lombard speech than for normal speech, albeit from a high baseline.

It is worth noting that while no net benefit of Lombard speech is seen in quiet for non-native listeners, and that the size of any such benefit for natives is likely to be limited, there may be individual modifications which are beneficial but whose effect is cancelled out by other changes which reduce intelligibility. Lombard speech induced by moderate levels of noise was neither beneficial nor detrimental overall, suggesting that some of the changes caused by higher noise levels may be responsible for the overall drop in keyword scores for Lombard96 relative to both normal and Lombard82 speech. We return to this point in Sec. IV.

Overall, the results of experiment I suggest that the observed Lombard advantage in noise has a dominant auditory basis which benefits both native and non-native listeners by substantial amounts. This interpretation is supported by the finding reported in Lu and Cooke (2009) that much of the Lombard benefit appears to stem from differences between the spectral profiles of Lombard and normal speech (see Fig. 2) which result in a greater release from energetic masking for the former.

The notion that energetic masking plays a key role in the Lombard advantage raises the question of whether the masker used in experiment I was better matched to normal rather than Lombard speech. Since that masker had the long-term average spectrum (LTAS) of normal speech, it is possible that some or all of the Lombard benefit was due to a mismatch between the LTAS of Lombard and normal speech. A second experiment measured the extent to which a match or mismatch between the LTAS of the target and masking speech influenced the Lombard benefit. Experiment II also tested the possibility that the choice of sentence material might have influenced the outcome of experiment I. While the Grid Corpus from which sentence material was derived is expected to have a high degree of uniformity, it is possible that the utterance subsets selected at random in experiment I may have differed in their intrinsic intelligibility. To eliminate the possibility of differential intelligibilities across sentence subsets, this factor was balanced across listeners in experiment II.

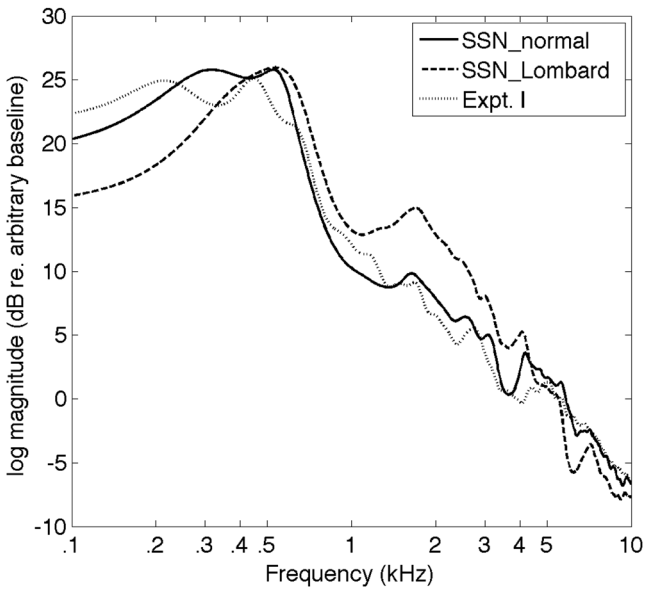

FIG. 2. Log magnitude spectra for the speech-shaped noise maskers used in experiments I and II. In each case, spectra were calculated using a 50-pole linear predictor fit to a 30 s segment of noise. Differences between the spectra SSN_normal and experiment I are due to the use of different subsets of speech material.

III. EXPERIMENT II: ROLE OF MASKER SPECTRUM AND SENTENCE SUBSET

A. Listeners

A fresh cohort of 93 listeners (73 females, 20 males; age 18–26, mean: 19.3) took part in experiment II. Listeners had a similar profile to those of experiment I. However, these students did the experiment at the beginning of their second year at university and consequently had approximately 8 months less exposure to English lectures and English language courses than those who participated in experiment I. Fifteen listeners were subsequently excluded from the analysis for the following reasons. One listener reported hearing problems, while two listeners had a first language other than Spanish or Basque. Four participants did not complete the experiment, and eight listeners were statistical outliers (scores $< 76\%$ in one of the two quiet conditions). The analysis reported here is based on the remaining 78 listeners.

B. Speech material and maskers

Utterances were drawn from the same pool as used for the "normal" and "Lombard96" conditions of experiment I. However, while experiment I used 50 stimuli selected at random in each condition, here eight utterances from each of eight talkers were used, leading to 64 tokens per condition. Unlike experiment I, which used a single speech-shaped noise masker whose long-term spectrum equaled that of the entire Grid Corpus, experiment II employed two new maskers, denoted "SSN_normal" and "SSN_Lombard," which were constructed using the normal and Lombard96 speech material respectively. Speech-shaped noise maskers were generated by filtering a 30 s sequence of uniformly distributed random numbers through a filter based on a 50-pole fit to the long-term spectrum of the normal or Lombard speech material used in experiment II. Log magnitude spectra of SSN_normal and SSN_Lombard are shown in Fig. 2, together with the spectrum of the masker used in experiment I. Compared to the spectrum of normal speech, Lombard speech exhibits two characteristic features of noise-induced speech: an upwards shift in the energy peak below 1 kHz caused by the combined effects of increases in F0 and F1, and a reduced spectral tilt (see also Stanton *et al.*, 1988; Sluijter and van Heuven, 1996). The second factor results in significantly more energy in the 1–3 kHz region relative to normal speech.

Experiment II contained 6 conditions resulting from the combination of two speech types (normal and Lombard96) and three presentation conditions (quiet, SSN_normal, SSN_Lombard). To eliminate speech subset choice effects, 6 sets of sentences were chosen and combined with maskers to

produce 36 sets of stimuli. Stimulus sets were then assigned to conditions using a 6×6 Latin square design. Participants were assigned to one of the six sets of stimuli in a balanced fashion. For the four masked conditions, normal and Lombard speech were combined with one of the two maskers at an SNR of -9 dB.

C. Results

Prior to analyzing the main factors of speech and masker type, any potential effect of the set of sentences heard in each condition was assessed. An analysis of variance (ANOVA) with two within-subject factors (masker condition: quiet, SSN_normal, SSN_Lombard; speech type: normal, Lombard96) and one between-subject factor (speech subset) showed no significant overall effect of subset [$p = 0.78$]. Data from the six subsets were combined for subsequent analyses.

Compared to experiment I, listeners' scores in quiet were lower in experiment II (7% versus 10.4% errors for normal speech). In noise (Fig. 3) scores were also substantially lower than in the equivalent masker conditions of experiment I (SSN_normal). While some differences between the masker spectra for experiment I and SSN_normal are apparent, they are small and unlikely to account for the large difference in scores, and in any case cannot explain the differences observed in quiet. Instead, differences in the amount of English exposure and phonetic training between the cohorts are a more probable cause. The experiment I cohort undertook the experiment at a later stage of their taught degree. Scores for both groups prior to phonetic training on an independent test of English intervocalic consonant identification in quiet and speech-shaped noise (test sets 1 and 4 of Cooke *et al.*, 2010) also indicated substantial differences: the cohort of experiment I had mean scores of 82% and 57% in quiet and noise respectively compared to 77% and 50% for the cohort of experiment II. The difference

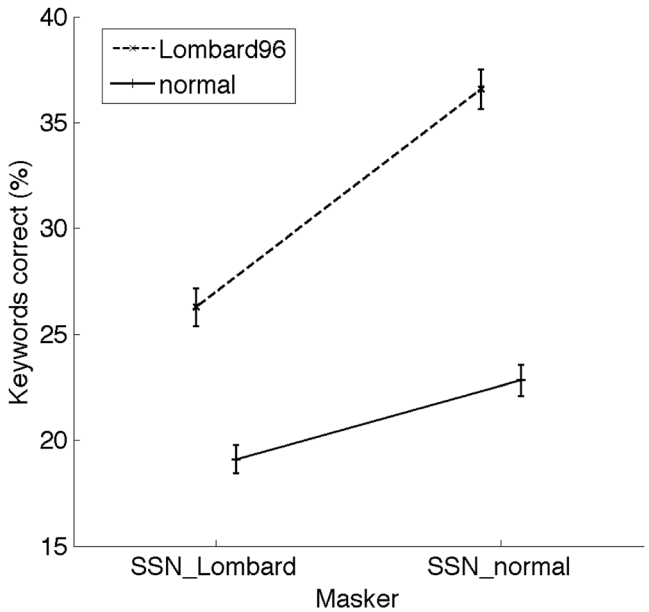

FIG. 3. Mean keywords correctly identified by non-native listeners in normal and Lombard speech in speech-shaped noise maskers derived from normal and Lombard speech.

between the groups would be amplified by the fact that the listeners undertook experiment I after approximately 2 months of phonetic training while the experiment II cohort performed the experiment prior to phonetic training.

In quiet, listeners identified 89.6% of the keywords correctly for normal speech and 87.4% for Lombard96 speech [$t(77) = 3.72, p < 0.001$]. Notwithstanding cohort differences, the percentage points benefit for normal speech is almost identical to that observed in experiment I with different listeners and different subsets of Grid sentences, confirming that Lombard speech induced by high levels of noise is slightly less intelligible than normal speech for non-native listeners when presented in quiet.

The masker based on Lombard speech resulted in significantly lower scores than the masker derived from normal speech. This was true for both normal and Lombard96 target sentences: for Lombard96 utterances, the Lombard_SSN masker reduced scores by more than 10 percentage points [$t(77) = 11.5, p < 0.001$], while for normal utterances the Lombard-based masker caused a drop of nearly 4 percentage points [$t(77) = 4.9, p < 0.001$]. This finding demonstrates that differences in the long-term spectral profile of normal and Lombard speech play an important role in the extent to which target utterances are masked.

However, the Lombard benefit was maintained regardless of masker spectrum. For the masker derived from normal speech, the Lombard advantage of around 14 percentage points was somewhat lower than the 19 percentage points seen in the equivalent condition in experiment I, but given the lower baseline for normal speech, this score actually represents a larger relative improvement (59% versus 51%). A smaller Lombard benefit of around 8 percentage points was observed in masking noise derived from Lombard speech. A repeated-measures ANOVA over the two types of SSN confirmed the interaction between SSN type and speech material [$F(1, 77) = 24.0, p < 0.001, \eta^2 = 0.039$].

Lombard speech studies typically show wide inter-speaker variability in the size of noise-induced effects (Stanton *et al.*, 1988; Junqua, 1993). Figure 4 plots keyword scores for normal and Lombard96 speech from each of the eight talkers, both in quiet and combined across the two masker types. Clear differences in intelligibility across talkers of up to 15 percentage points in quiet and up to 30 in noise can be seen. In quiet, the disadvantage for speech produced in noise is not universal. However, 4 of the 8 talkers show a gain for normal speech while for two of the remaining talkers (1 and 8) scores are probably near ceiling for non-native listeners. In noise, the Lombard benefit is apparent for every talker, although the size of the benefit varies from 5 to 18 percentage points. Further, talkers who have a high intrinsic intelligibility (as indicated by listener scores in quiet) are also more intelligible in noise [normal: $r = 0.82, p < 0.05$; Lombard96: $r = 0.89, p < 0.01$].

D. Interim discussion

Experiment II replicates and extends the principal findings of experiment I, eliminating a possible influence of choice of speech material. The Lombard benefit observed in

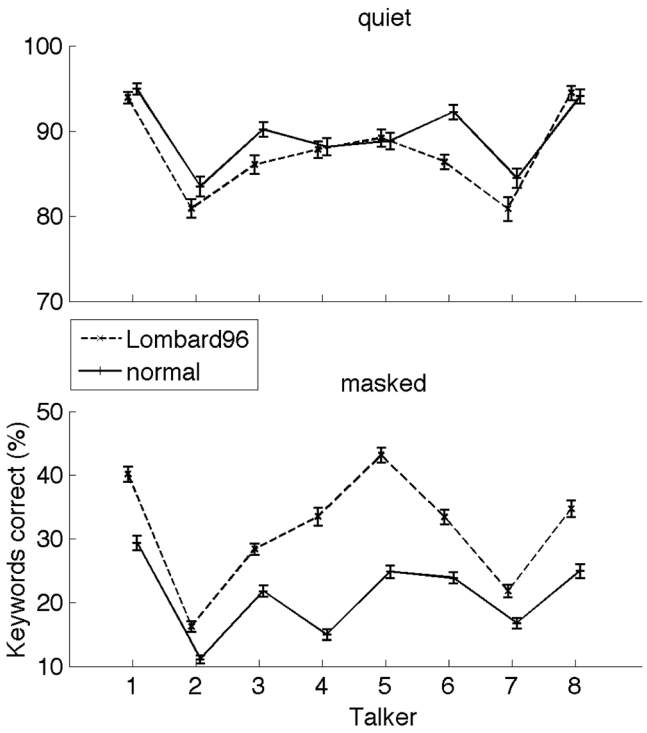

FIG. 4. Keyword identification scores for individual talkers in quiet (upper) and combined across the two maskers (lower).

noise was reversed in quiet conditions. The effect was present for speech material from half the talkers, and in no case was Lombard speech from a talker significantly more intelligible than normal speech from that talker.

Experiment II also clarifies the role of the masker spectrum in determining the extent of Lombard advantage seen in noise. A matched masker-target (e.g., normal speech masked by SSN based on normal speech) did not automatically result in lower scores: normal speech targets were more adversely affected by the Lombard masker. Instead, the Lombard masker lowered scores for both types of speech target relative to the normal speech masker. We interpret this result as demonstrating that intelligibility is a function of both the likelihood that any given spectral region is masked (a likelihood which is maximized in the case of a matching masker), and also the importance of that frequency region to speech perception, as, for example, reflected in the band importance functions of objective intelligibility measures such as SII (ANSI, 1997) (see also Gilbert and Micheyl, 2005; Ma *et al.*, 2009). The Lombard masker has around 5 dB more masking potential in the region centered on 2 kHz which is where the second and third formants of speech are typically encountered. The observation that a masker based on Lombard speech is more potent than one formed from normal speech is, of course, not unrelated to the hypothesis that much of the Lombard intelligibility benefit is derived from energetic masking release in those same regions which convey important speech information.

Lombard speech was less intelligible in quiet by the same amount in experiments I and II in spite of differences in the absolute levels of scores. We interpret the score differences from the two cohorts as resulting from differing levels of phonetic competence, and it is noteworthy that even the group who had received several months of perceptual training in English sounds found Lombard speech harmful relative to normal speech.

IV. GENERAL DISCUSSION

A talker's speech is affected by a range of contextual factors—acoustic environment, task, instructions, interlocutor—but the purpose and overall effect of the modifications can differ in each case. While a talker may intentionally clarify his or her speech when explicitly asked to, or when talking to a listener who is perceived as having comprehension difficulties, or when talking to an audience, there is no guarantee that such clarifications occur in response to noise, or even that a speaker has total control over their modified speech. Indeed, it seems probable that noise-induced modifications are in part designed to help overcome the effect of noise and are not conceived to be beneficial in quiet conditions, and may even be harmful, just as shouted speech lowers intelligibility in noise (Pickett, 1956; Rostolland, 1985). The outcome of the current study supports the view that the intelligibility benefits of Lombard speech result from low-level auditory factors which promote the audibility of target speech energy in frequency regions of high importance for speech perception, and not from intrinsically clearer speech, for instance, of the type which results from an instruction to speak clearly.

Both experiments found a consistent decrease in intelligibility for Lombard speech presented in quiet, albeit by a small amount, and it is worth asking what might be responsible for the fall in keyword scores. One possibility is that some of the parameters which convey phonological distinctions are also changed when speech is produced in the presence of noise. One example is the use of durational information to signal phonological voicing (i.e., the shortening of vowels preceding fortis consonants and the relatively longer duration of fortis vs lenis consonants in English). Since segment durations are modified in Lombard speech (Summers *et al.*, 1988; Junqua, 1993), it is conceivable that there are negative interactions with durational cues to voicing. Sankowska *et al.* (2011) found that while vowel shortening in the environment of voiceless codas was still present in Lombard speech, the durational cue for following-consonant voicing was significantly reduced relative to adult-directed and foreigner-directed speech. Since Lombard speech contains other segment-specific modifications such as a shift in energy from consonants to vowels (Junqua, 1993; Womack and Hansen, 1996), it seems reasonable to expect other potentially negative interactions with phonological cues to stem from noise-induced speech.

However, it would be premature to conclude from the results in quiet that linguistic enhancements are not present in Lombard speech. First, since we measure only the net benefit of one speech style over another, it is possible that linguistic enhancements do exist in Lombard speech but are outweighed by others (such as vowel lengthening and consonant shortening) which reduce intelligibility in quiet conditions. In noise, the latter may be rendered less salient via masking. Second, since the current study involved non-native listeners,

it is necessary to consider possible language-specific aspects, both in terms of non-native sensitivity to and weighting of modifications made to an L2, and also regarding non-native expectations about Lombard speech modifications which presumably stem mainly from their L1 experience.

Cross-linguistic research has shown (Bradlow and Bent, 2002; Smiljanic and Bradlow, 2009) that clear speech has features (such as vowel space expansion and decreases in speech rate) which are found cross-linguistically, whereas other parameters (e.g., long/short vowel contrasts, fortis/lenis stop contrasts) are implemented differently according to the phonologically relevant cues in a particular language. This mix of language-specific and cross-linguistic features could account for the fact that in the above studies non-native listeners did benefit from clear speech (presumably due to the presence of cross-language features and a certain level of L2 learning) but less than native listeners. Non-natives listeners may lack sufficient knowledge to take advantage of L2-specific cues, or those cues may conflict with their L1-based expectations or with their L2-interlanguage system. For example, a vowel duration increase implemented to enhance the fortis/lenis consonantal contrast in English might lead non-natives to assume that the vowel is tense rather than lax, while ignoring the consonantal contrast it was intended to convey, since vowel duration is often the main cue used by English L2 learners for tense/lax vowel contrasts (Bohn, 1995). This interpretation is consistent with the studies on cross-language clear speech mentioned above, which hypothesize that more experienced non-native listeners benefit more from clear speech because of their greater acquaintance with the language-specific features in the L2 (Smiljanic and Bradlow, 2011).

Similarly, Lombard speech may be more intelligible in noise because of phonetic-phonological changes that increase distinctiveness. In this case, too, some of the modifications—such as vowel space changes (Garnier, 2007)—may be cross-linguistic, in which case it is expected that experience of L1 Lombard speech could transfer to the L2 and result in intelligibility benefits there too. The finding that Spanish Lombard speech shows similar tendencies as those found in English (Castellanos *et al.*, 1996) is relevant for the listener group of the current study. However, there may also be language-specific modifications, such as the differing use of VOT in different languages (Smiljanic and Bradlow, 2009). Further cross-language comparisons of Lombard speech are required to address this possibility.

Lu and Cooke (2009) found that around a third of the intelligibility benefit of Lombard speech in noise was not accounted for by spectral modifications. If, as the current study suggests, linguistic enhancements play a minor role in the intelligibility advantage of Lombard speech, the question remains as to what factors underlie the benefit. Lu and Cooke (2009) speculated that some or all of the remaining intelligibility gain might be due to the slower speech rate of Lombard speech. For example, the mean sentence duration for the normal condition was 1.58 s compared to 1.76 s in the Lombard96 condition (here, sentence length is a proxy for the inverse of speech rate since all sentences have the same number of words). Evidence for a beneficial effect of a slower speech rate is mixed, with some studies finding increased intelligibility in noise (Cox *et al.*, 1987; Jones *et al.*, 2007; Bond and Moore, 1994; Hazan and Markham, 2004) and others revealing no advantage (Sommers, 1997; Uchanski *et al.*, 2002).

Determining which factors underlie perceptual benefits of modified speech styles and understanding how any advantages scale with background noise intensity or interact with a listener's native language is a key step in developing speech-generation algorithms capable of promoting intelligibility in applications of speech output technology. Text-to-speech systems in particular provide the scope for intelligibility-enhancing intervention at a range of points in the generation sequence, ranging from syntactic or lexical simplification through pause insertion and repetition at the suprasegmental levels, to instantaneous changes in spectral profile. Further studies which manipulate individual parameters such as speech rate at both global and segmental scales are needed to better understand the origin of intelligibility benefits of altered speech styles such as Lombard and clear speech.

V. CONCLUSIONS

In the presence of a masker, non-native listeners identified more keywords in Lombard sentences than in normal sentences. However, in quiet conditions Lombard speech was somewhat less intelligible than normal speech. Taken together with previous estimates of the Lombard advantage for native listeners in noise, which show that native listeners benefit only slightly more than non-native listeners from Lombard speech, these findings argue for a limited role for factors other than energetic masking release in explaining the increased intelligibility of Lombard speech. Both normal and Lombard sentences were more effectively masked by noise whose long-term spectrum matched that of Lombard speech, suggesting that Lombard speech places energy in parts of the spectrum of importance for speech perception, a characteristic which is beneficial when Lombard speech is the target, and harmful when Lombard speech is the masker.

ACKNOWLEDGMENTS

This work was partially funded by the Listening Talker (LISTA) project, supported by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number 256230. We also received support from the Spanish Ministerio de Ciencia e Innovación, Grant No. FFI2009-10264/FILO and award IT311-10 from the Gobierno Vasco. We thank Vincent Aubanel and Vasilis Karaiskos for help in running the native listener test at the University of Edinburgh.

APPENDIX A: THE INTELLIGIBILITY OF NORMAL AND LOMBARD SPEECH IN QUIET FOR NATIVE LISTENERS

Native listeners identified keywords in quiet from sentences in the normal and Lombard96 conditions used in experiment I. Thirty-four young adult listeners recruited

from the student population at the University of Edinburgh took part in the experiment. All had British English as their native language. Listeners were paid for their participation. Following audiometric screening, results from six listeners were excluded. Stimuli were drawn from the same sets of speech material used in experiment I but one difference here was that each of the two conditions consisted of 120 sentences, 15 from each of the 8 talkers. To exclude the possibility of intelligibility differences related to the choice of sentences in the two conditions, two sets of stimuli were constructed using the same utterances spoken in normal or Lombard conditions. Half of the listeners heard each subset. Listeners were tested individually in sound-attenuating booths using Beyerdynamic DT770 Pro headphones and the same custom program as used in experiment I. Following a short practice session, participants responded to the two blocked conditions, which were presented in a balanced order across listeners.

As expected, native listeners made very few errors in keyword identification for both types of speech material presented in quiet. Nevertheless, Lombard speech was less well identified than normal speech (1.45% versus 0.9% errors, corresponding to 97 and 60 keywords incorrect across the listener groups). While these differences are small, a Wilcoxon signed rank test conducted on rationalized arcsin-transformed scores indicated that the benefit for normal speech over Lombard speech was statistically significant $[V = 157, p = 0.012]$. Response times, measured from the end of the stimulus to the point at which the second of the two keywords was input, were slightly longer for Lombard speech (849 vs 835 ms) but the difference was not statistically significant $[p = 0.15]$.

[1]Having Spanish or Basque as a first language was the inclusion criterion. In practice, all were native Spanish speakers with different degrees of competence in Basque. The phonological systems of Spanish and Basque are very similar at a segmental level. There is only one English consonant, /ʃ/, which is present in Basque and absent in Spanish as a phoneme but even those participants who do not speak Basque are exposed to it on a regular basis through common usage words such as "kaixo" /kaiʃo/, which means "hello."

[2]In Lu and Cooke (2008) these conditions were denoted "Ninf_82" and "Ninf_96."

ANSI. (**1997**). S3.5-1997, *American National Standard Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

Black, J., and Hast, M. (**1962**). "Speech reception with altering signal," J. Speech Hearing Res. **5**, 70–75.

Bohn, O.-S. (**1995**). "Cross-language perception in adults: First language transfer doesn't tell it all," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York Press, Baltimore), pp. 379–410.

Bond, Z., Moore, T., and Gable, B. (**1989**). "Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask," J. Acoust. Soc. Am. **85**, 907–912.

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

Bořil, H. (**2008**). "Robust speech recognition: analysis and equalization of Lombard effect in Czech corpora," Ph.D. thesis, Czech Technical University, Prague.

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

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech i: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Castellanos, A., Benedí, J.-M., and Casacuberta, F. (**1996**). "An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect," Speech Commun. **20**, 23–35.

Chen, F. (**1980**). "Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level," Master's thesis, Massachusetts Institute of Technology, Cambridge.

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

Cooke, M., Barker, J., Cunningham, S., and Shao, X. (**2006**). "An audio-visual corpus for speech perception and automatic speech recognition," J. Acoust. Soc. Am. **120**, 2421–2424.

Cooke, M., García Lecumberri, M. L., and Barker, J. (**2008**). "The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception," J. Acoust. Soc. Am. **123**, 414–427.

Cooke, M., García Lecumberri, M. L., Scharenborg, O., and Van Dommelen, W. (**2010**). "Language-independent processing in speech perception: identification of English intervocalic consonants by speakers of eight European languages," Speech Commun. **52**, 954–967.

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**, 2059–2069.

Cox, R., Alexander, G., and Gilmore, C. (**1987**). "Intelligibility of average talkers in typical listening environments," J. Acoust. Soc. Am. **81**, 1598–1608.

Cutler, A., and Butterfield, S. (**1990**). "Durational cues to word boundaries in clear speech," Speech Commun. **9**, 485–495.

Cutler, A., Weber, A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

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

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. (**2007**). "Communiquer en environnement bruyant: De l'adaptation jusqu'au forcage vocal (Communication in noisy environments: From adaptation to vocal straining)," Ph.D. thesis, Universite Paris 6.

Gat, I., and Keith, R. (**1978**). "An effect of linguistic experience: auditory word discrimination by native and non-native speakers of English," Audiology **17**, 339–345.

Gilbert, G., and Micheyl, C. (**2005**). "Influence of competing multi-talker babble on frequency-importance functions for speech measured using a correlational approach," Acta Acust. Acust. **91**, 145–154.

Hansen, J. H. L. (**1996**). "Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition," Speech Commun. **20**, 151–170.

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

Jones, C., Berry, L., and Stevens, C. (**2007**). "Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners," Comp. Speech Lang. **21**, 641–651.

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

Li, C.-N. (**2004**). "Intelligibility of non-native lombard speech for non-native listeners," J. Acoust. Soc. Am. **115**, 2393–2394.

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

Lu, Y., and Cooke, M. (**2008**). "Speech production modifications produced by competing talkers, babble and stationary noise," J. Acoust. Soc. Am. **124**, 3261–3275.

Lu, Y., and Cooke, M. (**2009**). "The contribution of changes in f0 and spectral tilt to increased intelligibility of speech produced in noise," Speech Commun. **51**, 1253–1262.

Ma, J., Hu, Y., and Loizou, P. (**2009**). "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," J. Acoust. Soc. Am. **125**, 3387–3405.

Meador, D., Flege, J., and Mackay, I. (**2000**). "Factors affecting the recognition of words in second language," Bilingualism: Lang. Cognit. **3**, 55–67.

Payton, K. L., Uchanski, R. M., and Braida, 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., Durlach, N., and Braida, L. (**1985**). "Speaking clearly for the hard of hearing. I. Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **28**, 96–103.

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

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

Rostolland, D. (**1985**). "Intelligibility of shouted speech," Acustica **57**, 104–121.

Sankowska, J., García Lecumberri, M. L., and Cooke, M. (**2011**). "Interaction of intrinsic vowel and consonant durational correlates with foreigner directed speech," Poznan Stud. Contemp. Ling. **47**, 109–119.

Sluijter, A., and van Heuven, V. (**1996**). "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **100**, 2471–2485.

Smiljanic, R., and Bradlow, A. (**2011**). "Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness," J. Acoust. Soc. Am. **130**, 4020–4032.

Smiljanic, R., and Bradlow, A. R. (**2009**). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," Lang. Linguistics Compass **3**, 236–264.

Sommers, M. (**1997**). "Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment," J. Acoust. Soc. Am. **101**, 2278–2288.

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

Studebaker, G. (**1985**). "A rationalized arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Summers, W., Pisoni, D., Bernacki, R., Pedlow, R., and Stokes, M. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analysis," J. Acoust. Soc. Am. **84**, 917–928.

Uchanski, R., Geers, A., and Protopapas, A. (**2002**). "Intelligibility of modified speech for young listeners with normal and impaired hearing," J. Speech Lang. Hear. Res. **45**, 1027–1038.

Uchanski, R. M. (**2005**). "Clear speech," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell Press, Oxford, UK), Chap. 9, pp. 207–235.

Womack, B., and Hansen, J. (**1996**). "Classification of speech under stress using target driven features," Speech Commun. **20**, 131–150.