Spectral and temporal changes to speech produced in the presence of energetic and informational maskers[a]

Martin Cooke[b]
Ikerbasque (Basque Science Foundation) and Language and Speech Laboratory, Facultad de Letras, Universidad del Pais Vasco, Paseo de la Universidad 5, Vitoria, Alava 01006, Spain

Youyi Lu
Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, United Kingdom

(Received 3 December 2009; revised 5 May 2010; accepted 15 July 2010)

Talkers change the way they speak in noisy conditions. For energetic maskers, speech production changes are relatively well-understood, but less is known about how informational maskers such as competing speech affect speech production. The current study examines the effect of energetic and informational maskers on speech production by talkers speaking alone or in pairs. Talkers produced speech in quiet and in backgrounds of speech-shaped noise, speech-modulated noise, and competing speech. Relative to quiet, speech output level and fundamental frequency increased and spectral tilt flattened in proportion to the energetic masking capacity of the background. In response to modulated backgrounds, talkers were able to reduce substantially the degree of temporal overlap with the noise, with greater reduction for the competing speech background. Reduction in foreground-background overlap can be expected to lead to a release from both energetic and informational masking for listeners. Passive changes in speech rate, mean pause length or pause distribution cannot explain the overlap reduction, which appears instead to result from a purposeful process of listening while speaking. Talkers appear to monitor the background and exploit upcoming pauses, a strategy which is particularly effective for backgrounds containing intelligible speech.
© *2010 Acoustical Society of America.* [DOI: 10.1121/1.3478775]

PACS number(s): 43.70.Mn, 43.72.Dv, 43.66.Dc [MSS] Pages: 2059–2069

I. INTRODUCTION

It is a common experience to hold conversations in less-than-quiet conditions. Background noise has long been known to affect speech production (Lombard, 1911), with subsequent studies (e.g., Dreher and O'Neill, 1957; Charlip and Burk, 1969; Pisoni *et al.*, 1985; Summers *et al.*, 1988; Junqua, 1993; Letowski *et al.*, 1993; Tartter *et al.*, 1993; Pittman and Wiley, 2001; Hansen and Varadarajan, 2009) elaborating some of the main consequences, which include increases in speech level, fundamental frequency and vowel duration. However, the impact of different types of noise backgrounds, especially those containing intelligible speech, has received less attention.

In perception, a distinction is drawn between the energetic and informational masking effects of noise on speech. Energetic masking (EM) refers to the reduction in intelligibility of a target source caused by spectro-temporal overlap with a background source at the level of the auditory periphery, while informational masking (IM) describes any additional negative consequences for intelligibility produced by more central processes. For instance, informational masking covers the inability to segregate target and background components even though they are audible (Brungart, 2001), or the effect of the extra cognitive load associated with processing more than one simultaneous sound source (Mattys *et al.*, 2009). Informational masking is especially important when the interfering sound source is a competing talker (Carhart *et al.*, 1969; Brungart, 2001).

The current study was motivated by the question of whether the differing energetic and informational masking potential of background sources leads to different effects on speech *production*. By analogy with the effect of maskers on speech perception, it is possible that strong informational maskers such as competing speech have an impact over and above the effect of energetic maskers. For example, intelligible speech in the background might interfere with a talker's ability to monitor an interlocutor, or to turn-take correctly. Talkers might adopt different speaking strategies in the face of informational maskers.

Classic Lombard speech studies have not addressed the issues raised by informational masking on speech production. Of the few studies that have examined the effect of background speech on a foreground speech production task, Webster and Klumpp (1962) used talker-listener pairs seated face to face and communicating word lists in conditions of quiet and ambient noise. When there was one background talker-listener pair, the speech level of the foreground talker increased by up to 9 dB compared to the condition without the background pair. The speaking rate in words per second decreased slightly when the background pair was present and

---

[a] A preliminary report of this work appeared at Interspeech 2009 in Brighton, United Kingdom.

[b] Author to whom correspondence should be addressed. Electronic mail: m.cooke@ikerbasque.org

the foreground pair made more communication errors when talking at the same time as the competing pair.

More recently, Lu and Cooke (2008) measured speech produced in N-talker babble, for a range of values of N from 1 (single competing talker) to ∞ (speech-shaped stationary noise), a continuum designed to produce varying amounts of energetic and informational masking (Simpson and Cooke, 2005). Talkers read sentence prompts in N-talker babble equalized for presentation level. The competing speech masker led to smaller speech production modifications (in F0, speech level, spectral center of gravity and duration) than stationary noise, with intermediate changes seen for other values of N, suggesting that, at least for these parameters, the degree of speech modification is largely correlated with the energetic masking potential of the noise. Lu and Cooke did find some effects—more false starts and an increased number of short pauses—of a competing speech masker beyond those expected on the basis of energetic masking and which might indicate the impact of informational masking. However, the effects were small, and no evidence was found of speaking strategies which exploited the temporal fluctuations of competing utterances.

The near absence of effects related to the informational masking potential of the background in Lu and Cooke (2008) might have been due to both the unnatural and the non-interactive nature of the task. Like many Lombard studies, talkers read a list of words/sentences without receiving feedback about the success or failure of their communication. Consequently, there was little incentive for the talkers to consciously change their speech even with masking noise present in the headphones. As Lane and Tranel (1971) pointed out, talkers do not necessarily make changes in order to communicate better with themselves. Summers *et al.* (1988) also suggested that much larger changes might have been observed in the acoustic-phonetic properties of the utterances produced under masking noise if some form of communication was provided to the talkers. Various studies examined the effect of noise on a communicative task, using talker-listener pairs to establish a conversation (Korn, 1954; Gardner, 1964), to communicate word/utterance lists (Webster and Klumpp, 1962; Bořil, 2008) and to complete interactive cooperative tasks in an instructor-follower manner (Rivers and Rastatter, 1985; Mixdorff *et al.*, 2007; Patel and Schell, 2008). In these studies, increases in F0, speech level and word duration in noise conditions compared to quiet were reported, consistent with non-communicative tasks. Other studies evaluated the effect of the presence or absence of a communicative task on noise-induced speech production modifications (Junqua *et al.*, 1998; Junqua *et al.*, 1999; Garnier, 2007). Junqua and colleagues compared speech produced when reading a list of phrases with that produced while talking to a voice dialing system. In both tasks the background was quiet or contained wideband noise. Noise-induced speech modifications in both tasks were similar: increases in utterance intensity, vowel F0 and duration of sonorant phonemes over those produced in quiet. However, relative to the reading task, and irrespective of the presence of noise, the communication factor led to decreases in utterance intensity and phoneme duration, and increases in F0. More recently, Garnier (2007, ch. 4) found larger Lombard effects when talkers completed an interactive task with a partner than when carrying out a non-interactive task alone.

These studies of the effect of communicative task and noise on speech production used energetic maskers. The current study examines the impact of both energetic and informational maskers on speech production using communicative and non-communicative tasks designed to be moderately natural. In one condition, we used competing speech backgrounds since these produce strong informational masking effects (Carhart *et al.*, 1969; Brungart, 2001, Cooke *et al.*, 2008). We were interested in whether speech production is affected by speech maskers in ways which parallel speech perception. When presented at the same level as stationary speech shaped noise, competing speech is a far less effective energetic masker (Festen and Plomp, 1990), presumably because listeners exploit temporal and spectral dips in the masker to glimpse the target speech. On the other hand, the greater informational masking effect of speech in the presence of speech might force talkers to adopt other strategies to maintain intelligibility in the face of a potentially-confusable background. One research question is whether talkers are able to retime their contributions to exploit glimpses, benefiting listeners sharing the same noise environment. A second question concerns the communicative environment: are speech production changes induced by the background source stronger in a task involving real communication?

To explore these issues, we contrasted two task conditions using the same puzzle-solving task: speaking aloud, or while communicating with a partner who was also subject to the same noise backgrounds. We carried out several types of analysis on speech produced in quiet and three noise backgrounds chosen for their energetic and/or informational masking capacity. First, we employed standard Lombard speech metrics (changes in mean F0, intensity, spectral tilt and duration) to check whether our speech elicitation procedure produced similar task and noise type effects seen in previous studies (Garnier, 2007; Lu and Cooke, 2008). To address the question of whether talkers exploited masker modulations to reduce energetic masking for the listener, we explored the temporal overlap between speech and background. Finally, we examined several measures of vowel space dispersion known to reflect intelligibility (Bradlow *et al.*, 1996) to investigate whether talkers responded differentially to speech and non-speech backgrounds.

II. SPEECH COLLECTION

A. Talkers

Eight native speakers of British English (4 males and 4 females) drawn from staff and students in the Department of Computer Science at the University of Sheffield participated in the corpus collection. All received a hearing test using a calibrated software audiometer which was used to test each ear separately at 250, 500, 1000, 2000, 4000, and 8000 Hz. All had normal hearing level. Ethics permission was obtained following the University of Sheffield ethics procedure. Participants were grouped into 4 pairs, and in each both members were of the same gender.

TABLE I. Noise and task conditions in each recording session.

Session 1	Participant 1 of pair condition: Q task: No communication	Both participants condition: Q task: Communication	Participant 2 of pair condition: Q task: No communication
Session 2	Conditions: SMN, SSN, CS balanced across participants task: No communication		
Session 3	Conditions: SMN, SSN, CS balanced across participants task: Communication		

B. Background noise types

In addition to collecting speech in quiet (Q) conditions, we employed three types of noise: competing speech (CS), speech-shaped noise (SSN) and speech-modulated noise (SMN). SMN is produced by modulating speech-shaped noise with the short-term temporal envelope of speech, and has approximately the same EM potential as natural speech but lacks intelligibility and thus is devoid of IM (Festen and Plomp, 1990). Its use here allows the additional IM effect of natural speech to be distinguished from the EM produced by speech. Speech-shaped noise and speech-modulated noise are pure energetic maskers, but differ in that SMN contains temporal modulations which talkers might be able to exploit to reduce the effect of EM.

Maskers were based on speech produced in quiet conditions from two male and two female talkers (session 1; see Table I and Sec. II D below). Ten minutes of speech from each of the four talkers was transcribed manually using Wavesurfer v1.8.4[1] to identify speech/nonspeech segments and silent pauses. Sound types such as *uh, um, ooh*, paper-rustling, breathing, laughing, coughing, and unintelligible utterances were labeled as *nonspeech*. Silent pauses longer than 100 ms were also marked. Each *nonspeech* segment was replaced with a silence of the same duration. The resulting four signals were used as the competing speech maskers. Note that no participant heard his or her own voice in subsequent conditions. For each competing speech masker, the corresponding speech-shaped noise was generated by filtering white noise with a filter whose spectrum equaled the long-term spectrum of the speech segments of the competing masker, and the corresponding speech-modulated noise was produced by modulating the generated speech-shaped noise with the envelope of the competing speech masker using a procedure described in Brungart (2001).

C. Task

Talkers were asked to solve Sudoku puzzles either alone or in pairs. Sudoku puzzles are widely-practised and quite compelling for many people, and constitute an interesting task which can be extended over a reasonable duration by presentation of new puzzles. Speech produced in these conditions is moderately-natural while participants are involved in the task. One advantage of Sudoku puzzles is that they elicit many repetitions of spoken digits. We used these items as robust ‘anchor’ tokens for comparison of spectral and durational parameters across conditions. Puzzles of intermediate difficulty were selected from the Daily Sudoku website,[2] the level chosen on the basis of pilot tests which demonstrated that easy puzzles led to less need for communication, while more difficult exemplars produced longer pauses and less interaction.

In all conditions, participants were instructed to solve as many puzzles as possible. When solving puzzles individually, participants were asked to speak aloud to describe the progress of puzzle-solving. In the condition where participants worked in pairs, a visual barrier separated the talkers, and both members of the pair were subject to the same background conditions. Unlike in the study of Webster and Klumpp (1962) where talkers were instructed to maintain a high level of communicative accuracy, no such instructions were given here on the basis that a focus on accuracy might detract from the naturalness of the task. In practice, speakers interacted fluidly in cooperative puzzle solving and reported that the task was engrossing.

D. Procedure

Participants produced speech in the 8 conditions derived from all combinations of task (alone/in pairs) and background (Q, SSN, CS, SMN). Each participant attended three recording sessions with tasks and noise conditions allocated as shown in Table I. Speech material was recorded in quiet in the first session, with speakers solving puzzles alone or with a partner. Subsequent sessions involved speaking in the presence of noise, either speaking alone (session 2) or communicating with a partner (session 3). Identical maskers were used for corresponding noise conditions of sessions 2 and 3. Condition order was balanced across participants. Recordings lasted 10 min for each noise condition. In each of the three sessions, participants were given several Sudoku puzzles and asked to keep solving alone or together up to the 10 min time limit. In practice, most individuals or pairs worked on a single puzzle in the 10 min recording period. Participants were permitted a short break between conditions. No talker heard his or her own voice in the competing speech maskers since the presence of competing speech from the same talker is known to produce more informational masking for listeners (Brungart, 2001; Cooke *et al.*, 2008; Vestergaard *et al.*, 2009).

E. Recording setup

Corpus collection sessions took place in an IAC single-walled acoustically-isolated booth, with a table placed inside. When working together, talkers sat at opposite sides of the table which had a thick cardboard screen in the middle in

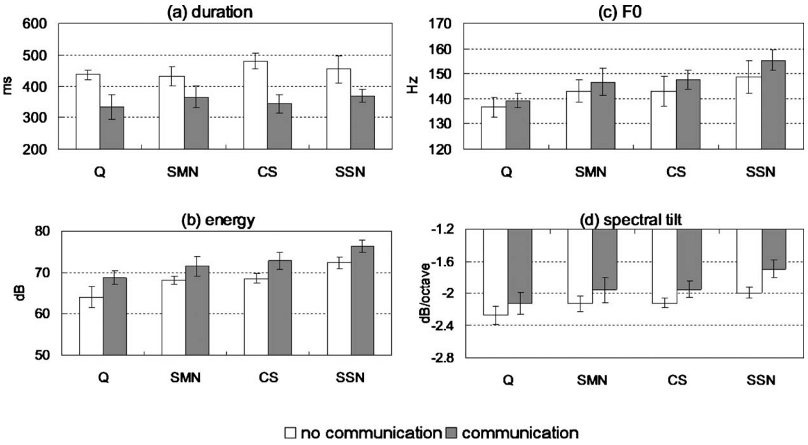


FIG. 1. Acoustic parameters of speech produced as a function of background and task. Values shown are means over talkers and error bars, here and elsewhere, indicate 95% confidence intervals. Background conditions are quiet (Q), speech-modulated noise (SMN), competing speech (CS) and stationary speech-shaped noise (SSN).

order to provide some acoustic isolation to reduce crosstalk as well as to require the talkers to rely only on acoustic cues to decode each other's speech. Two Bruel & Kjaer (B & K) type 4190 1/2 in. microphones each coupled with a preamplifier (B & K type 2669) were fixed on the screen and directed toward each talker. When seated, the distance between the talker and the nearest microphone was set at approximately 20 cm.

Recorded signals were passed to a conditioning amplifier (B & K Nexus model 2690) prior to digitalization at 25 kHz with a Tucker-Davis Technologies (TDT) System 3 RP2.1. In the noise conditions, maskers were presented diotically over Sennheiser HD 250 Linear II headphones using the same TDT system. A presentation level of 82 dB SPL was used since this within the range known to provide sufficient EM [Summers *et al.* (1988) used 80, 90, and 100 dB, Junqua (1993) and Garnier (2007) used 85 dB, while Pittman and Wiley (2001) used 80 dB] but still relatively low in order to elicit IM effects. A more intense competing speech background could cause a release from IM during the foreground talker's speech production if listeners were able to exploit level differences to separate their own speech and that of the background (Brungart, 2001; Cooke *et al.*, 2008). When solving puzzles alone, participants sat at one side of the table with no other differences to the setup. Signal collection and masker presentation was under computer control.

Participants wore the headphones throughout, including for the quiet condition, to ensure that own-voice masking was held at a constant level. In order to compensate the sound attenuation introduced by the closed ear headphones, each participant's own voice was fed back via the TDT system and mixed with the noise signal prior to presentation over the headphones. At the beginning of each recording session, each talker was asked to talk freely to the microphone while wearing the headphones. The level of voice feedback was manually and iteratively adjusted until the talker felt the overall loudness level matched that when not wearing the headphones. Voice feedback level was then held constant for the whole recording session and participants were unable to adjust the level.

F. Transcription

Boundaries of the digits "one" to "nine" were marked manually, as were speech/nonspeech segments and silent pauses of length greater than 100 ms, a value chosen to avoid closures associated with stop consonants. Although significant crosstalk (approximately 12 dB level difference between the two microphones) was present when both talkers were speaking simultaneously, this did not affect the ability to locate speech segment boundaries or speech/nonspeech segments. Fewer than 5% of digit words appeared in the overlapped sections. These were omitted from the analysis. On average, 12.3 clear instances of each digit (s.d. 4.2) were available per talker in each condition. Subsequent analyses were based on the transcribed digits and the speech/nonspeech timings.

III. RESULTS

A. Lombard effects

Measures of word duration, root-mean-square energy and mean fundamental frequency (F0) were obtained for each transcribed digit using PRAAT v4.3.24 (Boersma and Weenink, 2005). F0 estimates were provided at 10 ms intervals using an autocorrelation-based method (Boersma, 1993) implemented in the PRAAT program. Spectral tilt estimates came from a linear regression of the long-term average spectrum (0–8 kHz), expressed in dB/octave, implemented in MATLAB. To avoid undue influence from outliers, medians were used to estimate per-talker values for each acoustic parameter in each condition.

Figure 1 displays mean parameter estimates across talkers. Lombard effects compatible with those of previous studies can be seen, with clear increases in most parameter values in noise relative to quiet. The presence of a communicative intent also led to significant changes in most parameters relative to the task with no communication. However, unlike Garnier (2007) (Chap. 4), who found stronger Lombard effects in a communicative task, here the factors of task and background did not interact for any of the measured parameters. That is, relative to the quiet baseline, the increases or decreases induced by noise were no different across the two types of task.

Separate two-way repeated measure ANOVAs with within-subjects factors of task and background (2 tasks $\times 4$ backgrounds) for each of the 4 parameters supported visual impressions.[3] In both tasks, compared to quiet, SSN produced the largest increases and the other two maskers,

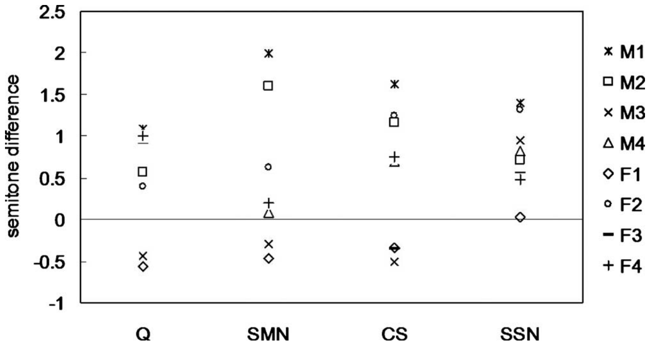


FIG. 2. Mean individual talker F0 differences between communicative and non-communicative tasks for each background.

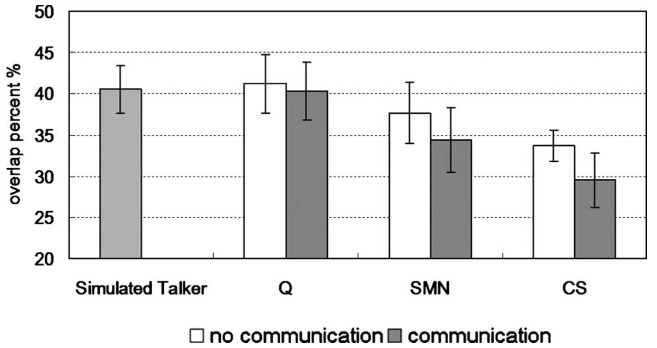


FIG. 3. Foreground-background overlap percentage as a function of task and background. Values shown are averages across the 8 talkers. The left-most bar shows the degree of overlap for simulated talkers described in Sec. III B 4.

SMN and CS, led to comparable but smaller changes in energy, mean F0 and spectral tilt. This was confirmed by the significant main effect of type of background [$F(3,21)=17.98$, $p<0.001$, $\eta^2=0.72$ for energy; $F(3,21)=8.98$, $p<0.01$, $\eta^2=0.56$ for F0; $F(3,21)=7.70$, $p<0.01$, $\eta^2=0.52$ for spectral tilt]. For these 3 parameters, post-hoc pairwise comparisons (here and elsewhere by paired *t*-tests with Bonferroni-adjustment) between the 4 types of background collapsed across tasks showed that the differences between SMN and CS were not significant, but that both differed from quiet and SSN ($p<0.05$). Word duration was similar and statistically identical across the 4 types of background in both tasks.

Figure 1 also shows a clear task effect, with larger values for F0 and intensity, a flattening of spectral tilt, and shorter duration for the collaborative task. The main effect of task type was significant for duration [$F(1,7)=29.16$, $p<0.01$, $\eta^2=0.81$], energy [$F(1,7)=26.08$, $p<0.01$, $\eta^2=0.79$] and spectral tilt [$F(1,7)=28.57$, $p<0.01$, $\eta^2=0.80$]. A task effect for these 3 acoustic parameters was further confirmed by post-hoc pairwise comparisons in each background condition ($p<0.05$), apart from spectral tilt in which the difference in competing speech condition was not significant ($p=0.12$). Mean F0 also tended to increase when the communication factor was present across all 4 background conditions (Fig. 1) although only the task effect in speech-shaped noise background was significant ($p<0.05$) and the main effect of task type on F0 also failed to reach statistical significance [$F(1,7)=3.74$, $p=0.07$]. Further inspection of the task effect for individual talkers showed a clear trend for increased F0 across tasks for most participants and background conditions (Fig. 2). However, certain talkers producing a decrease in F0 in some conditions, highlighting the impact of individual variation in Lombard speech.

The presence of standard Lombard effects confirms that the tasks and relatively low noise levels employed in the current study are suitable for inducing noise-related speech production changes, and the size of the changes in parameter values for different noise types supports the finding in our earlier study (Lu and Cooke, 2008) with read speech that it is largely the energetic masking potential of the noise that determines the amount by which speech production is modified. The two maskers designed to produce similar amounts of EM, namely the competing speech and speech modulated noise conditions, do indeed lead to similar sized Lombard effects.

B. Temporal modifications

1. Foreground-background overlap

Here, the issue of whether talkers could avoid overlapping in time with a noise background was studied by measuring the degree of temporal overlap between speech activity in the foreground talker and speech or "speech-like" activity (in the case of SMN) in the background masker. Overlap values were computed relative to the length of speech from the foreground, expressed as overlap percentage, in order to normalize for differences in the amount of speech produced across conditions. For each talker, the overlap was computed between the foreground speech segments produced in the backgrounds with temporal fluctuations (i.e., competing speech or speech-modulated noise) and the background in which the speech was collected, shown as "CS" and "SMN" in Fig. 3. While there is of course no overlap for speech produced in quiet, a reference overlap value for such speech can be computed using the background used in the fluctuating masker case (shown as "Q" in Fig. 3). If talkers were attempting to make use of the gaps in the fluctuating background, one would expect to see a smaller degree of overlap relative to the quiet reference in which the foreground speech is independent of the background.

Compared to quiet, both tasks produced a reduction in overlap in the conditions of speech modulated noise and competing speech maskers, with a greater reduction for the latter. The difference between backgrounds was statistically significant [$F(2,14)=44.82$, $p<0.001$, $\eta^2=0.87$]. Further, for both tasks, the differences between individual conditions of quiet and SMN, and between SMN and CS were also significant ($p<0.01$). Background noise type and task did not interact ($p=0.11$). However, post-hoc pairwise comparisons reported that compared to the task with no communication, the communicative task led to a significantly smaller overlap percent in the backgrounds of SMN ($p<0.05$) and CS ($p<0.01$) but produced statistically-identical values in quiet ($p=0.25$). The task effects in SMN and CS conditions also resulted in a significant main effect of task type [$F(1,7)=110.39$, $p<0.001$, $\eta^2=0.94$]. Finding the same

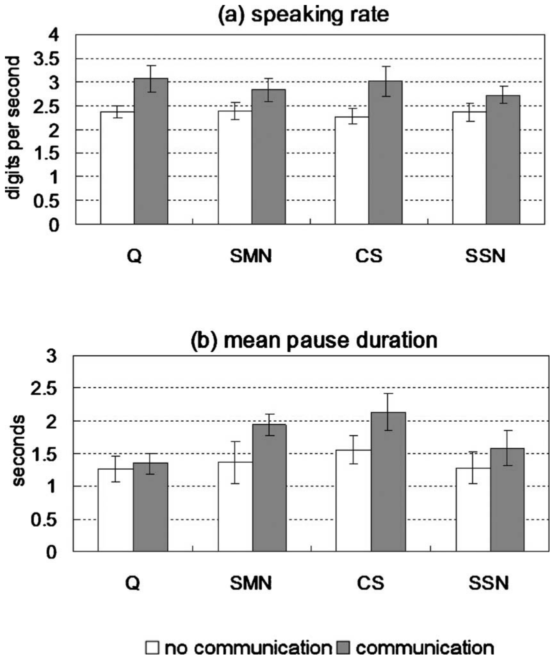


FIG. 4. (a) Speech rate and (b) mean pause duration for speech produced as a function of task and background.

overlap value in the two quiet conditions is unsurprising because there is no reason to expect talkers to retime their contributions in a noise-free background.

On the surface, these results suggest that talkers are capable of retiming their contributions in a way which reduces overlap with a modulated masker in the background. However, there are a number of ways in which talker might reduce foreground-background overlap in modulated noise. Talking more rapidly or changing the distribution of pause lengths could conceivably result in overlap reduction *without any active attempt* to retime contributions relative to the background. Subsequent analyses explored this issue.

2. Speech rate

The mean speech rate in each condition and for each talker was estimated using the digits extracted during corpus transcription. To accommodate the different number of exemplars of each digit in each condition, a specific number n_i of each of the digits $i=1..9$ (different for each digit but fixed across conditions) was chosen, and speaking rate $rate_c$ for condition c was computed according to:

$$rate_c = \frac{\sum_{i=1}^{9} n_i}{\sum_{i=1}^{9} \sum_{k=1}^{n_i} d_{cik}},$$

where d_{cik} is the duration of the kth exemplar of digit i in condition c.

The upper panel of Fig. 4 shows across-talker means of speech rate by task and background type. A clear increase in speaking rate for the communicative task relative to the task without communication can be seen [$F(1,7)=23.12$, $p<0.01$, $\eta^2=0.77$]. This task effect is present in all background conditions ($p<0.05$). While the difference in speaking rate across tasks as shown in Fig. 4 might at first sight be considered as a contributory factor given the task differences in SMN and CS conditions observed in Fig. 3, this is unlikely since in the quiet condition there was no task effect on overlap yet the task produced a significantly faster speaking rate. Further, the effect of noise background was not significant ($p=0.29$) and none of the speaking rate differences between background conditions reached significance ($p>0.05$). It appears that speech rate changes cannot account for the overlap reduction either as a function of task type or noise background.

3. Mean pause duration

Another factor which could lead to reduced overlap is a change in pause structure as a function of the background or task. Mean pause durations shown in the lower panel of Fig. 4 do indeed show both task and background effects. The communicative task resulted in longer pauses overall [$F(1,7)=9.70$, $p<0.05$, $\eta^2=0.58$], although not in quiet. Both tasks showed longer pauses in the modulated noise conditions. For the communicative task, this trend was statistically significant [$F(1.98,13.88)=9.04$, $p<0.01$, $\eta^2=0.56$]. Comparison of Figs. 3 and 4 reveals a common pattern: longer pause durations correlate strongly with decreasing amounts of overlap ($r=-0.90$, $p<0.05$). This finding is consistent with the idea that talkers wait until an appropriate point to make their contributions in the face of a modulated background. Alternatively, it might be thought that the mere presence of noise results in longer pauses. However, the data of Fig. 4 suggest otherwise: the mean pause duration for stationary noise is barely different from quiet ($p>0.05$). While speech-shaped noise produced the largest Lombard effects (Fig. 1), it has little effect on pause duration.

4. Simulated talkers

There remains the possibility that the distribution of pause duration varies as a function of the background (e.g., participants matching their rhythm to that of a competing talker) without necessarily requiring *active* timing of contributions to avoid overlap. Example distributions of pause and contribution lengths for a single talker in quiet, modulated noise and competing speech background are shown in Fig. 5. Clear differences between the distributions include a greater number of long pauses in the competing speech condition and very frequent short pauses in the modulated noise background. To accommodate the long one-sided tail, these distributions can be modeled using gamma distributions (Abramowitz and Stegun, 1972) whose density is given by

$$f(x;\alpha;\beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)},$$

with parameters α ("shape") and β^{-1} ("rate"), and where $\Gamma(\alpha)$ is the gamma function. Fits and associated parameters are shown in Fig. 5.

To test the idea that the distribution of pauses might explain the observed overlap percentage in each condition, simulated talkers were constructed, with each "talker" consisting simply of sequences of numbers defining durations of pauses and speech contributions, drawn from distributions of pauses and contributions of speech produced in each condi-

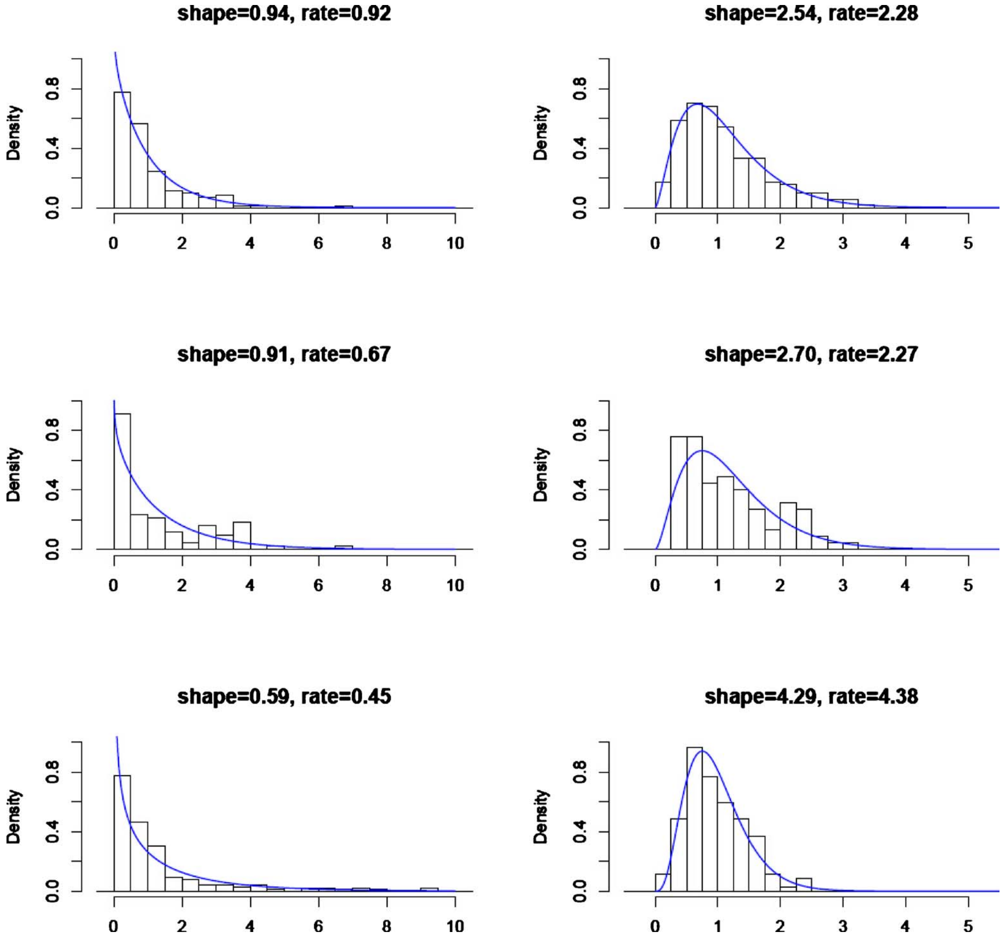


FIG. 5. (Color online) Example pause length (left) and contribution length (right) densities for a single talker in quiet (top), speech modulated noise (middle) and a competing talker background (bottom). The horizontal axis shows duration in seconds. Gamma density fits are also plotted with shape and rate values shown.

tion of the experiment. A talker's pause structure in each condition was then simulated by alternately sampling from the gamma distributions for pauses and contributions to produce a sequence of the same length as the original talker data. One hundred simulation sequences were produced for each condition.

The overlap rates for the simulated talkers were statistically-identical across tasks and noise backgrounds. The degree of overlap for the simulated talkers is plotted as the leftmost bar in Fig. 3 and matches very closely the real talker data in the quiet condition. An additional simulation was performed by permuting consecutive pause-contribution pairs from the original data (i.e., without the gamma distribution fits). Again, overlap scores (40%) similar to those in quiet were obtained. These simulations demonstrate that random sampling from the different pause and contribution duration distributions cannot account for the differences in overlap rate across tasks and backgrounds.

Taken together, the findings reported in this section provide strong support for the idea that talkers are capable of timing their speech in ways which reduce temporal overlap with a modulated background. The effect is present when talking alone but is stronger when involved in a communicative task. Talkers were able to exploit both modulated backgrounds tested here, but showed better overlap avoidance for intelligible speech compared to modulated noise.

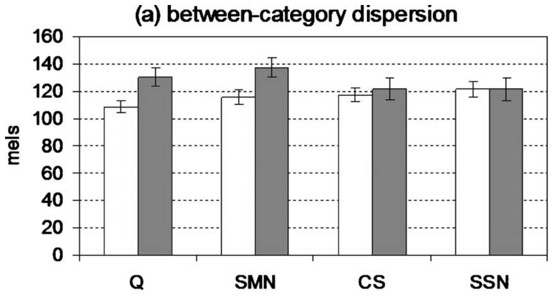
C. Vowel space dispersion

Vowel space dispersion is known to affect speech intelligibility (Bradlow *et al.*, 1996) and studies by Bond *et al.* (1989), Garnier (2007, ch. 5), and Bořil (2008) hinted at the presence of differences in vowel space dispersion for speech produced in noise relative to quiet, but these were not examined statistically. To explore a possible effect of task and noise types relative to quiet on expansion or compactness of the F1-F2 vowel space, F1 and F2 frequencies were estimated for the manually-segmented vowels /iː/, /ɪ/, /e/ and /uː/ in the words "three," "six," "seven" and "two" respectively in the 8 conditions (2 tasks × 4 backgrounds). Formant frequencies were the average of the central 3 frames in each vowel instance and computed using the Burg algorithm (Burg, 1975) implemented in PRAAT. F1 and F2 values were then converted to the perceptually-motivated mel scale (Fant, 1973).

$$M = (1000/\log_{10} 2) \times \log_{10}((F/1000) + 1),$$

where M and F are frequencies in mels and Hertz respectively.

To provide a single quantitative indicator of vowel space expansion or compactness between vowel categories for each of the 8 conditions, a measure of each talker's "between-category dispersion" was calculated as the mean of the Eu-



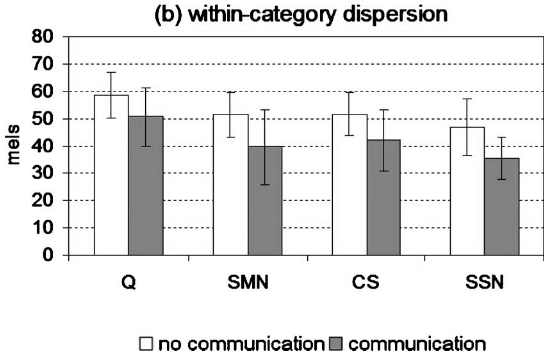


FIG. 6. (a) Between-category dispersion and (b) within-category dispersion for a subset of vowels produced in the 4 types of background (quiet and 3 noisy backgrounds) with or without a communication factor. Values shown are averages across the 8 talkers.

clidean distances of each vowel token from the central point in that talker's F1-F2 space. A second measure of the compactness of individual vowel categories, "within-category dispersion," was also obtained. First, the mean of the Euclidean distances of each individual vowel token from the category mean was computed, as for the measure of between-category dispersion. Then a single measure for each talker was calculated as the mean within-category dispersion across all four vowel categories. These metrics follow Bradlow *et al.* (1996).

Figure 6 (upper panel) shows that, compared to the task with no communication, the communicative task led to larger between-category dispersion in the quiet ($p < 0.01$) and speech-modulated noise ($p < 0.001$) conditions, but produced similar values in the other two conditions, suggesting that the effect of the communication factor on between-category dispersion differed with background type, confirmed by a significant interaction between task and background [$F(1.51, 10.53) = 6.71$, $p < 0.05$, $\eta^2 = 0.49$]. Compared to quiet, none of the 3 noise background led to significant changes in between-category dispersion for the communicative task. Although there was a tendency to see increased dispersion for the 3 noise backgrounds relative to quiet in the speaking aloud task, only SSN differed from the quiet condition ($p < 0.05$).

Figure 6 also shows the within-category dispersion measure (lower panel). Here, there was no significant interaction between task and background [$F(3, 21) = 0.05$, $p = 0.91$]. For both tasks, the three noise backgrounds led to decreases in within-category dispersion—tighter clustering of exemplars for each vowel category—[$F(3, 21) = 9.13$, $p < 0.05$, $\eta^2 = 0.62$], with the largest fall in the SSN condition. Within-category dispersion was significantly smaller in the communicative task [$F(1, 7) = 10.35$, $p < 0.05$, $\eta^2 = 0.59$] although post-hoc comparisons between tasks in each of the 4 background conditions showed that only the difference between tasks in speech-shaped noise condition reached significance ($p < 0.05$).

Summarizing, the effect of noise and task on between-category dispersion was mixed, with the communicative task leading to more contrast in the vowel space for two of the four backgrounds. A more consistent picture is seen for within-category dispersion, which was reduced by independent effects of task and noise.

IV. DISCUSSION

A. Energetic and informational masking effects of background noise on speech production

All noise backgrounds led to speech production changes observed in most Lombard studies: increases in speech level and F0 and a flattening of spectral tilt (more energy at higher frequencies). The size of these modifications was comparable for the two temporally-modulated backgrounds (SMN and CS), and larger for speech-shaped noise (SSN). When matched for level SSN is a more effective energetic masker (EM) than SMN and CS (Festen and Plomp, 1990), and since SMN and CS were designed to produced similar amounts of EM, this outcome supports our hypothesis (Lu and Cooke, 2008) that the scale of Lombard effects is proportional to the energetic masking potential of the background. In response to noise, an increase in speech level benefits speech intelligibility due to an increase in signal-to-noise ratio, as well as the flattening of spectral slope which enables more of the speech to escape masking, at least for the maskers used here which had a low-frequency bias. By contrast, increases in F0 might be secondary effects correlated with a change in speech level (Schulman, 1985; Gramming *et al.*, 1988; Stevens, 2000), and have been found to contribute little to speech intelligibility in noise (Bond and Moore, 1994; Barker and Cooke, 2007; Lu and Cooke, 2009) or in quiet (Bradlow *et al.*, 1996; Assmann *et al.*, 2002). Indeed, some authors argue that the greater sparsity of harmonics resulting from F0 increases might even reduce intelligibility (Chladkova *et al.*, 2009) due to the undersampling hypothesis (Ryalls and Lieberman, 1982), which claims that greater between-category vowel dispersion is necessary to offset the reduced information conveyed by (fewer) harmonics.

While changes in level and spectral properties of noise-induced speech are well-known, the current study revealed for the first time the presence of clear temporal effects on speech production in the face of modulated maskers. Talkers attempted to avoid overlapping with fluctuating noise backgrounds in tasks with and without a communication factor, with a greater reduction for competing speech than for speech modulated noise, and even larger effects for the communicative task. Subsequent analyses suggested that this reduction in overlap cannot be accounted for by factors such as passive changes in speech rate or pause distribution, or indeed by a general noise effect on mean pause duration.

Avoidance of temporal overlapping of foreground speech with competing talker masker or speech-modulated noise leads to a release from EM by the background due to fewer foreground speech signal elements being obscured by the background noise, aiding segregation of foreground and background speech for the interlocutor. The additional overlap reduction produced by the competing talker background relative to the speech-modulated noise may also result in reduced IM due to improved foreground-background segregation (Kidd *et al.*, 1994). It is unclear what kind of perceptual processes are used to drive the reduction in overlap. One possibility is that intelligibility of the competing speech masker relative to the speech-modulated noise allows a better prediction of upcoming pauses. This strategy is supported by the data of Figs. 4 and 5: for the competing speech background, there is evidence that the increased mean pause duration is largely due to a greater number of long pauses, perhaps due to talkers' monitoring the background for a suitable place to interject.

Interestingly, the present study found a reduction in foreground-background speech overlap when a competing speech masker was present not only in the communicative task but also in the task with no communication. No such tendency was found in our earlier study using a non-communicative task (Lu and Cooke, 2008), a difference which might stem from the nature of the competing speech material employed in the two studies. While the current work used a long section of spontaneous speech material, in Lu and Cooke (2008) we used short utterances (less than 3 s in duration) with almost all short pauses less than 100 ms. For short duration utterances, talkers may have been unable to attend to and track competing speech material sufficiently rapidly to modify their own productions in response.

While the current results showed the possible presence of a temporal-domain strategy to yield a release from energetic and informational masking, there are other mechanisms open to talkers. For example, differences in speech level or F0 between foreground and background are known to reduce IM (Bird and Darwin, 1998; Brungart, 2001; Vestergaard *et al.*, 2009). In the present study, observed changes in speech level and F0 in the competing speech condition appeared to be governed primarily by EM factors since speech-modulated noise induced very similar speech level and F0 changes. It may be that temporal domain speech manipulation is an efficient form of talker behavior compared to manipulations of vocal level and F0: increasing speech level is energy consuming and the extent to which talkers can manipulate F0 is constrained by physiological and articulatory constraints.

Another approach talkers might adopt to reduce informational masking effects is to ensure that foreground speech categories are both consistently produced within-category and well-separated across categories. In particular, vowel space expansion (i.e., greater discrimination between vowel categories) has been associated with an intelligibility advantage on the basis of intertalker differences in overall intelligibility within normal, conversational speech (Byrd, 1994; Bond and Moore, 1994; Bradlow *et al.*, 1996; Hazan and Markham, 2004) as well as on the basis of clear versus conversational style comparisons (Picheny *et al.*, 1986; Moon and Lindblom, 1994; Bradlow *et al.*, 2003; Smiljanić and Bradlow, 2005). Here, we found only a slight tendency toward vowel space expansion for the speaking alone task, and a stronger effect when talkers were engaged in the cooperative task, where expansion was observed for the quiet and speech-modulated noise conditions. Evidence from the literature for vowel space changes in response to noise is mixed. While Mixdorff *et al.* (2007) and Bořil (2008) reported similar tendencies to ours for vowel space expansion in Lombard speech, Bond *et al.* (1989) and Garnier (2007, ch. 5) found speech in noise to have a more compact vowel space than when produced in quiet. Some caution is needed in interpreting our own results, since they are based on only 4 vowels. While digit-based vowels were chosen as robust anchors for comparison across tasks and conditions, the 4 categories available lack two vertices of the vowel quadrilateral, so the room for expansion of the vowel space is limited. It is possible that greater vowel space expansion would be observed with a richer selection of vowel categories.

In contrast to the results for vowel space expansion, a clearer picture emerged of within-category clustering. In the presence of noise, or when communicating, talkers produced more consistent exemplars. This tendency of within-category vowel clustering due to a talker's more precise articulation of each vowel, also found in studies of clear speech (Chen, 1980), could benefit vowel discrimination because the more tightly clustered categories are less likely to lead to intercategory confusion, although Bradlow *et al.* (1996) showed that tightness of within-category clustering may not be a good correlate of perceptual performance.

It is unclear whether vowel space effects result from the EM or IM potential of a noise background. Competing speech did not produce stronger effects than pure energetic maskers, suggesting a limited role for IM. On the other hand, stationary noise, which produces greater EM than competing speech, did not exert a significantly larger effect on the vowel space measures than IM, so it is possible that it some combination of EM and IM potential is responsible for the observed changes in vowel space.

B. Task effects

Substantial task effects were observed throughout the present study for all background conditions, including quiet. Indeed, communication itself acts in some ways like noise in terms of increases in speech level and F0 and a flattening of spectral tilt. These changes confirm the findings of Garnier (2007), although for speech level are at odds with those of Junqua *et al.* (1998), who observed a level decrease when talkers were in communication with a speech recognition device compared to reading word lists. Bořil (2008) attributes the latter finding to participants consciously lowering their voices to obtain accurate results from the recognition system. The fact that in the current study changes in speech are produced by communication even in the quiet condition suggests that talkers imagine that an increase in speech level can help create more intelligible speech at the ears of the speech partner, while the observed flattening of spectral tilt places more speech energy in regions above 1 kHz known to be important for speech perception (French and Steinberg, 1947; Black, 1959; Schum *et al.*, 1991; Studebaker and Sherbecoe, 1991).

Word durations were substantially smaller in the communicative task, supporting Junqua *et al.* (1998). The reduction in word length of up to 25% might result from talkers' attempts to maintain a fluid interaction with their partner. Moderate increases in speech rate do not necessarily degrade intelligibility (Bradlow *et al.*, 1996; Uchanski *et al.*, 2002).

In the present study, there was no *additional* effect of communication on the size of the noise-induced speech production changes in speech level, F0 and spectral tilt. This is at odds with some previous studies which reported a larger speech production change from quiet to noisy condition due to the effect of communication. Garnier (2007, ch. 4; see also Garnier *et al.* 2010) observed a larger shift of spectral energy toward higher frequencies as well as greater increases in speech level and F0 from quiet to noisy conditions when subjects were interacting with a speech partner compared to while talking alone. A greater noise-induced effect in speech level when a communication factor is present is also reflected in the steeper slope of linear regression of vocal intensity as a function of ambient noise level for communicational speech (0.5, Webster and Klumpp, 1962; 0.39, Gardner, 1964; 0.29–0.61, Gardner, 1966) compared to read speech (0.11, Dreher and O'Neill, 1957; 0.12, Lane *et al.*, 1970; 0.15, Egan, 1972). Nevertheless, other studies reflect the pattern of the current study. For instance, comparing the size of Lombard effects between the tasks with and without a communication factor, Junqua *et al.* (1998) showed a similar degree of F0 change. For participants reading word/sentence lists, Kryter (1946) and Pickett (1958) also reported slopes of the voice-noise level function at 0.33 and 0.40 respectively, which are similar to values obtained for communicative tasks (0.39: Gardner, 1964; 0.29–0.61: Gardner, 1966). The difference between the current pattern and that in Garnier (2007, ch. 4) might be explained by the higher baseline in quiet for speech produced in the communicative task. Without noise exposure, Garnier (2007, ch. 4) found that the communication factor led to small speech production changes, while here we found significant changes in speech level and spectral tilt for the communicative task relative to speaking alone. Since a very forceful vocal effort may degrade the speech intelligibility due to distortion compared to normal speech production (Pickett, 1956; Rostolland, 1985), the discrepancy might also result from a ceiling effect especially in the most adverse condition (SSN) when talking to a speech partner. A further difference between the current study and Garnier (2007, ch. 4) concerns the availability of visual information: Garnier's participants were able to see each other while here they were not.

V. CONCLUSIONS

Noise backgrounds which differ in their energetic and informational masking potential might be expected to have different impacts on speech production. The current study found that while changes in speech level and a number of spectral properties scaled with the energetic masking capacity of the background, temporal changes of talkers' speech were strongest for the competing speech background where IM was dominant. While talkers were able to exploit masker modulations to retime their contributions even in the case of a pure energetic masker, their ability to avoid temporal overlap with the background was greatest for competing speech, suggesting that talkers monitor the background to a level sufficient to predict upcoming pauses. Overall, these findings suggest that when exposed to noise, talkers adopt a "listening-while-speaking" strategy which helps to increase the probability of message reception at the ears of the interlocutor. The benefit may arise both from a reduction in energetic masking, via spectral and temporal reallocation of speech energy to frequency regions and time intervals where it is least likely to be masked, and a partial release from informational masking, by affording easier separation of foreground and background speech for the interlocutor.

ACKNOWLEDGMENTS

Both authors are supported by the EU Marie Curie Research Training Network "Sound to Sense." We thank Mitch Sommers and the reviewers for helpful suggestions.

[1] http://www.speech.kth.se/wavesurfer (Last viewed 8/13/2010).

[2] http://www.dailysudoku.com/sudoku (Last viewed 8/13/2010).

[3] Unless otherwise stated, all subsequent statistical analyses followed this pattern i.e., 2-way repeated-measures ANOVA examining task and background effects.

Abramowitz, M., and Stegun, I. A. (**1972**). *Handbook of Mathematical Functions* (Dover, New York), Chap. 6.

Assmann, P. F., Nearey, T. M. and Scott, J. M. (**2002**). "Modeling the perception of frequency-shifted vowels," in International Conference on Spoken Language Processing, pp. 425–428.

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

Bird, J., and Darwin, C. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. Palmer, A. Rees, Q. Summerfield, and R. Meddis (Whurr, London).

Black, J. W. (**1959**). "Equally contributing frequency bands in intelligibility testing," J. Speech Hear. Res. **2**, 81–83.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in Proceedings of the Institute of Phonetic Sciences, Vol. **17**, pp. 97–110.

Boersma, P., and Weenink, D. (**2005**). "Praat: Doing phonetics by computer (version 4.3.14) [computer program]," http://www.praat.org.

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

Bond, Z. S., Moore, T. J., 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.

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

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

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.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Burg, J. P. (**1975**). "Maximum entropy spectrum analysis," Ph.D. thesis, Stanford University, Stanford, CA.

Byrd, D. (**1994**). “Relations of sex and dialect to reduction,” Speech Commun. **15**, 39–54.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). “Perceptual masking in multiple sound backgrounds,” J. Acoust. Soc. Am. **45**, 694–703.

Charlip, W. S., and Burk, K. W. (**1969**). “Effects of noise on selected speech parameters,” J. Commun. Disord. **2**, 212–219.

Chen, F. R. (**1980**). “Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level,” MS thesis, Massachusetts Institute of Technology, Cambridge, MA.

Chladkova, K., Boersma, P., and Podlipsky, V. J. (**2009**). “On-line formant shifting as a function of F0,” in Proceedings of Interspeech, Brighton, UK, pp. 464–467.

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

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

Egan, J. J. (**1972**). “Psychoacoustics of the Lombard voice response,” J. Aud Res. **12**, 318–324.

Fant, G. (**1973**). *Speech Sounds and Features* (MIT, Cambridge, MA).

Festen, J. M., and Plomp, R. (**1990**). “Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,” J. Acoust. Soc. Am. **88**, 1725–1736.

French, N. R., and Steinberg, J. C. (**1947**). “Factors governing the intelligibility of speech sounds,” J. Acoust. Soc. Am. **19**, 90–119.

Gardner, M. B. (**1964**). “Effect of noise on listening levels in conference telephony,” J. Acoust. Soc. Am. **36**, 2354–2362.

Gardner, M. B. (**1966**). “Effect of noise, system gain, and assigned task on talking levels in loudspeaker communication,” J. Acoust. Soc. Am. **40**, 955–965.

Garnier, M. (**2007**). “Communiquer en environnement bruyant: De l’adaptation jusqu’au forçage vocal (Communication in noisy environments: From adaptation to vocal straining),” Ph.D. thesis, l’Universite Paris 6, Paris, France.

Garnier, M., Henrich, N., and Dubois, D. (**2010**). “Influence of sound immersion and communicative interaction on the Lombard effect,” J. Speech Lang. Hear. Res. In press.

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

Hansen, J. H. L., and Varadarajan, V. (**2009**). “Analysis and compensation of Lombard speech across noise type and levels with application to in-set/out-of-set speaker recognition,” IEEE Trans. Audio, Speech, Lang. Process. **17**, 366–378.

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

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

Junqua, J. C., Fincke, S., and Field, K. (**1998**). “Influence of the speaking style and the noise spectral tilt on the Lombard reflex and automatic speech recognition,” in International Conference on Spoken Language Processing, pp. 467–470.

Junqua, J. C., Fincke, S., and Field, K. (**1999**). “The Lombard effect: A reflex to better communicate with others in noise,” in International Conference on Acoustics, Speech and Signal Processing, Vol. **4**, pp. 2083–2086.

Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). “Reducing informational masking by sound segregation,” J. Acoust. Soc. Am. **95**, 3475–3480.

Korn, T. S. (**1954**). “Effect of psychological feedback on conversational noise reduction in rooms,” J. Acoust. Soc. Am. **26**, 793–794.

Kryter, K. D. (**1946**). “Effects of ear protective devices on the intelligibility of speech in noise,” J. Acoust. Soc. Am. **18**, 413–417.

Lane, H. L., and Tranel, B. (**1971**). “The Lombard sign and the role of hearing in speech,” J. Speech Lang. Hear. Res. **14**, 677–709.

Lane, H. L., Tranel, B., and Sisson, C. (**1970**). “Regulation of voice communication by sensory dynamics,” J. Acoust. Soc. Am. **47**, 618–624.

Letowski, T., Frank, T., and Caravella, J. (**1993**). “Acoustical properties of speech produced in noise presented through supra-aural earphones,” Ear Hear. **14**, 332–338.

Lombard, E. (**1911**). “Le signe de l’elevation de la voix (The sign of the rise in the voice),” Ann. Maladiers Oreille, Larynx, Nez, Pharynx **37**, 101–119.

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

Lu, Y., and Cooke, M. P. (**2009**). “The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise,” Speech Commun. **51**, 1253–1262.

Mattys, S. L., Brooks, J., and Cooke, M. P. (**2009**). “Recognizing speech under a processing load: Dissociating energetic from informational factors,” Cognit. Psychol. **59**, 203–43.

Mixdorff, H., Pech, U., Davis, C., and Kim, J. (**2007**). “Map task dialogs in noise—A paradigm for examining Lombard speech,” in International Congress of Phonetic Sciences, pp. 1329–1332.

Moon, S.-J., and Lindblom, B. (**1994**). “Interaction between duration, context and speaking style in English stressed vowels,” J. Acoust. Soc. Am. **96**, 40–55.

Patel, R., and Schell, K. W. (**2008**). “The influence of linguistic content on the Lombard effect,” J. Speech Lang. Hear. Res. **51**, 209–220.

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.

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

Pickett, J. M. (**1958**). “Limits of direct speech communication in noise,” J. Acoust. Soc. Am. **30**, 278–281.

Pisoni, D. B., Bernacki, R. H., Nusbaum, H. C., and Yuchtman, M. (**1985**). “Some acoustic-phonetic correlates of speech produced in noise,” in International Conference on Acoustics, Speech and Signal Processing, pp. 1581–1584.

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

Rivers, C., and Rastatter, M. P. (**1985**). “The effects of multitalker and masker noise on fundamental frequency variability during spontaneous speech for children and adults,” J. Aud Res. **25**, 37–45.

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

Ryalls, J. H., and Lieberman, P. (**1982**). “Fundamental frequency and vowel perception,” J. Acoust. Soc. Am. **72**, 1631–1634.

Schulman, R. (**1985**). “Dynamic and perceptual constraints of loud speech (A),” J. Acoust. Soc. Am. **78**, S37.

Schum, D. J., Matthews, L. J., and Lee, F. (**1991**). “Actual and predicted word-recognition performance of elderly hearing-impaired listeners,” J. Speech Hear. Res. **34**, 636–642.

Simpson, S. A., and Cooke, M. P. (**2005**). “Consonant identification in *N*-talker babble is a nonmonotonic function of *N*,” J. Acoust. Soc. Am. **118**, 2775–2778.

Smiljanić, R., and Bradlow, A. R. (**2005**). “Production and perception of clear speech in Croatian and English,” J. Acoust. Soc. Am. **118**, 1677–1688.

Stevens, K. N. (**2000**). *Acoustic Phonetics* (MIT, Cambridge, MA).

Studebaker, G. A., and Sherbecoe, R. L. (**1991**). “Frequency-importance and transfer functions for recorded CID W-22 word lists,” J. Speech Hear. Res. **34**, 427–438.

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

Tartter, V. C., Gomes, H., and Litwin, E. (**1993**). “Some acoustic effects of listening to noise on speech production,” J. Acoust. Soc. Am. **94**, 2437–2440.

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

Vestergaard, M. D., Fyson, N. R. C., and Patterson, R. D. (**2009**). “The interaction of vocal characteristics and audibility in the recognition of concurrent syllables,” J. Acoust. Soc. Am. **125**, 1114–1124.

Webster, J. C., and Klumpp, R. G. (**1962**). “Effects of ambient noise and nearby talkers on a face-to-face communication task,” J. Acoust. Soc. Am. **34**, 936–941.