Influence of Sound Immersion and Communicative Interaction on the Lombard Effect

Maëva Garnier
Institut Jean Le Rond d'Alembert, LAM, (UMR 7190: UPMC Univ Paris 06, Centre National de la Recherche Scientifique (CNRS), Ministère de la Culture), Paris, France; and The University of New South Wales, Sydney, Australia

Nathalie Henrich
GIPSA Laboratory, Department of Speech and Cognition, Grenoble, France (UMR 5216: CNRS, Grenoble University)

Danièle Dubois
Institut Jean Le Rond d'Alembert, LAM (UMR 7190: UPMC Univ Paris 06, CNRS, Ministère de la Culture), Paris, France

Purpose: To examine the influence of sound immersion techniques and speech production tasks on speech adaptation in noise.
Method: In Experiment 1, we compared the modification of speakers' perception and speech production in noise when noise is played into headphones (with and without additional self-monitoring feedback) or over loudspeakers. We also examined how this sound immersion effect depends on noise type (broadband or cocktail party) and level (from 62 to 86dB SPL). In Experiment 2, we compared the modification of acoustic and lip articulatory parameters in noise when speakers interact or not with a speech partner.
Results: Speech modifications in noise were greater when cocktail party noise was played in headphones than over loudspeakers. Such an effect was less noticeable in broadband noise. Adding a self-monitoring feedback into headphones reduced this effect but did not completely compensate for it. Speech modifications in noise were greater in interactive situation and concerned parameters that may not be related to voice intensity.
Conclusions: The results support the idea that the Lombard effect is both a communicative adaptation and an automatic regulation of vocal intensity. The influence of auditory and communicative factors has some methodological implications on the choice of appropriate paradigms to study the Lombard effect.

KEY WORDS: Lombard effect, speech production, noise, headphones, communicative interaction

Speakers increase their vocal intensity when talking in noisy environments. This adaptation is called the *Lombard effect,* and it was originally interpreted as an automatic regulation of voice intensity from auditory feedback. After being reported qualitatively by Lombard (1911), this regulation effect was then quantified (Egan, 1972; Fairbanks, 1954; Lane, Tranel, & Sisson, 1970) and gave rise to many psychophysiological studies on the concept of the *audio-phonation loop*. Researchers have demonstrated how a similar regulation also occurs for other voice parameters, such as pitch (Elman, 1981, Ternström, Sundberg, & Collden, 1988) or spectral content (Burzynski & Starr, 1985, S. R. Garber, Siegel, & Pick, 1981), and how a perturbation of the audio-phonation loop can affect speech control and disfluency (Conture, 1974; S. F. Garber & Martin, 1977). Later, several studies provided arguments supporting the idea that this audio-phonation loop is underlined by a neural reflex (Bauer, Mittal, Larson, & Hain, 2006; Leydon, Bauer, & Larson, 2003; Nonaka, Takahashi, Enomoto, Katada, & Unno, 1997), or at least how its regulation is an uncontrollable behavior (Pick, Siegel, Fox, Garber, & Kearney,

1989), something that also has been observed in babies and animals (Amazi & Garber, 1982; Siegel, Pick, Olsen, & Sawin, 1976; Sinnott, Stebbins, & Moody, 1975).

Phonetic studies extended that original way of considering the Lombard effect, by showing how speech adaptation in noise consists not only of an increase of vocal intensity but also of global speech reorganization. Speech produced in noise (i.e., Lombard speech) has been characterized by an increase in intensity and pitch, a shift of spectral energy toward the medium frequencies, a decrease of speech rate, articulatory movements of greater amplitude, and phoneme modifications (Castellanos, Benedi, & Casacuberta, 1996; Davis, Kim, Grauwinkel, & Mixdorff, 2006; Garnier, 2008; Junqua, 1993; Kim, 2005; Mokbel, 1992; Stanton, Jamieson, & Allen, 1988; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). These speech modifications reduce to a large degree the efficiency of automatic speech recognition systems, which are usually based on models of conversational speech produced in quiet conditions (Hanson & Applebaum, 1990; Junqua, 1993). Conversely, Lombard speech is more intelligible for human listeners (Dreher & O'Neill, 1958; Pittman & Wiley, 2001; Van Summers et al., 1988), so some of its characteristics have been applied to speech enhancement techniques (Skowronski & Harris, 2006). This increased intelligibility of Lombard speech has led several authors to support the idea that speech adaptation in noise is also motivated by communication (Junqua, 1993; Lane & Tranel, 1971), similar to the adaptation observed in speech addressed to infants or people with hearing impairment (Lindblom, 1990; Lindblom, Brownlee, Davis, & Moon, 1992; Picheny, Durlach, & Braida, 1985, 1986). Providing support for that idea, Amazi and Garber (1982) and Junqua, Finckle, and Field (1999) showed how vocal intensity increases more in noise for conversational speech than in reading. Recent studies have brought to light the fact that speech modifications in noise are not only global over the whole utterance but also consist of the specific enhancement of some audiovisual cues to segment perception (Garnier, 2008) and of some prosodic cues to discourse structure and information enhancement (Garnier, Dohen, Lœvenbruck, Welby, & Bailly, 2006; Patel & Schell, 2008; Welby, 2006).

Concerns Raised by Experimental Paradigms

The extent to which this adaptation is an uncontrollable regulation of vocal intensity from the auditory feedback and/or a communicative effect that can be controlled individually is not very well understood yet, although it may be very important in the choice of an experimental paradigm. So far, the most classical paradigm to study Lombard speech has consisted of playing noise over headphones to individuals seated alone in a recording booth and reading out lists of words or sentences (Castellanos et al., 1996; Egan, 1972; Junqua, 1993; Stanton et al., 1988; Van Summers et al., 1988). However, this paradigm may raise a number of concerns.

First, closed headphones affect internal and external hearing, independently from any noise played into them. This may affect voice production as well as the perception of both one's own voice and that of the communication partner, in a manner similar to that which has been demonstrated for earplugs (Kryter, 1946; Tufts & Frank, 2003). Noisy environments may also be perceived differently when played over headphones, because the perception of soundscapes and the feeling of immersion in them depend on sound-restitution devices (Guastavino, Katz, Levitin, Polack, & Dubois, 2005). For all these reasons, the headphone paradigm may influence fundamentally the phenomenon of speech adaptation in noise.

A second concern is that the speech task has been shown to affect the increase of vocal intensity in noise that is greater in experimental situations where speakers have to search for intelligibility (Amazi & Garber, 1982; Junqua et al., 1999). Consequently, we can expect the modification of other speech parameters in noise to be influenced by the speech task, depending on the communicative involvement it requires.

Goals and Hypotheses of the Study

In this article, we discuss two experiments we conducted to test the influence of auditory and communicative factors on the Lombard effect and to determine the extent to which they may be neglected or must be taken into account in experimental protocols.

In Experiment 1, we explored whether using headphones to immerse a speaker in noise under laboratory conditions affects his or her speech adaptation from quiet to noisy conditions (i.e., the Lombard effect), in comparison to playing the noise over loudspeakers. Davis et al. (2006) already compared lip articulation under both paradigms and showed that headphones led to significantly more ample articulatory movements. Now, the improved quality of denoising techniques (Mixdorff, Grauwinkel, & Vainio, 2006; Ternström, Södersten, & Bohman, 2002) allows researchers to compare acoustic parameters of speech produced under both paradigms. Several studies have already examined acoustic aspects of Lombard speech, using a loudspeaker paradigm and a dedicated denoising technique (Södersten, Ternström, & Bohman, 2005; Ternstrom, Bohman, & Södersten, 2003, 2006; Ternström, Södersten, & Bohman, 2002). These three studies differed from previous studies not only because they used a different immersion method but also because

other aspects of their protocol were different (e.g., noise types, noise levels, measure of noise levels). Consequently, we cannot compare their results with previous studies using a headphones paradigm. We therefore decided to conduct a first experiment that was dedicated to the comparison of these two paradigms, for different levels of ambient noise, two different types of noise, and different acoustic parameters commonly investigated by studies of Lombard speech. We presumed that headphones may induce additional degradation of the auditory feedback, independently from the noise played into them, and may lead to a greater increase of speech parameters in noise than the loudspeaker paradigm.

To test that hypothesis further, we also investigated whether additional feedback of one's own voice in the headphones could compensate for the differences between headphones and loudspeakers paradigms. This hypothesis is supported by previous studies on the *sidetone effect,* which have shown how vocal intensity is decreased when speakers receive amplified feedback of their own voice (Lane, Catania, & Stevens, 1961; Lane et al., 1970). Kadiri (1998) also reported a reduced increase of F0 in noise when the speaker had additional feedback of his or her voice in the headphones, but no significant influence on the increase of syllable duration and first formant frequency in noise.

In Experiment 2, we explored whether speech adaptation in noise is affected when the task of speech production involves interacting with a speech partner. A number of previous studies have already included communicative aspects in their experimental task. Some of these required speakers to address read information to the experimenter (Davis et al., 2006) or to a machine interface (Junqua et al., 1999; Kim, 2005). In other studies, the speaker had feedback on the listener's comprehension by means of a vu-meter (Södersten et al., 2005; Ternström et al., 2006). An interactive game was used in a recent study on Lombard speech (Patel & Schell, 2008), and several earlier studies on the Lombard effect have also examined spontaneous speech (Gardner, 1966; Korn, 1954; Kryter, 1946). Although it is difficult to compare all of these previous studies because they have varying protocols (languages, type and level of noise, etc.), there is a tendency toward a greater increase of vocal intensity in noise for tasks that include communicative aspects. Indeed, the slope of the linear regression of vocal intensity as a function of ambient noise level ranged from 0.30 to 0.38 in studies on spontaneous speech (Gardner, 1966; Korn, 1954; Kryter, 1946) and only from 0.12 to 0.15 in studies on read speech (Egan, 1972; Lane et al., 1970). This observation was then confirmed by two studies dedicated to that comparison (Amazi & Garber, 1982; Junqua et al., 1999). The question of what happens to other speech parameters remains open. Experiment 2 was aimed at answering that question. Because of the strong correlation between vocal intensity, fundamental frequency, richness of the voice in high harmonics, first formant, and mouth opening (Schulman, 1989; Sundberg & Nordenberg, 2006; Titze, 1989), we expected these parameters to also increase more from quiet to noise in interactive conditions. Furthermore, if speech adaptation in noise consists not only of talking louder but also of speaking more clearly—by enhancing some segment and prosodic cues (Garnier, 2008; Garnier, Dohen, et al., 2006; Patel & Schell, 2008; Welby, 2006)—then we could expect such language-specific strategies to be adopted in interactive situations only.

Experiment 1: Comparison of Sound Immersion Techniques

Material and Methods

Participants and Protcol

This experiment was conducted with 10 native French speakers (5 men and 5 women) who ranged in age from 20 to 28 years. Only 1 had some basic knowledge about the Lombard effect. None of them presented any voice or auditory problems.

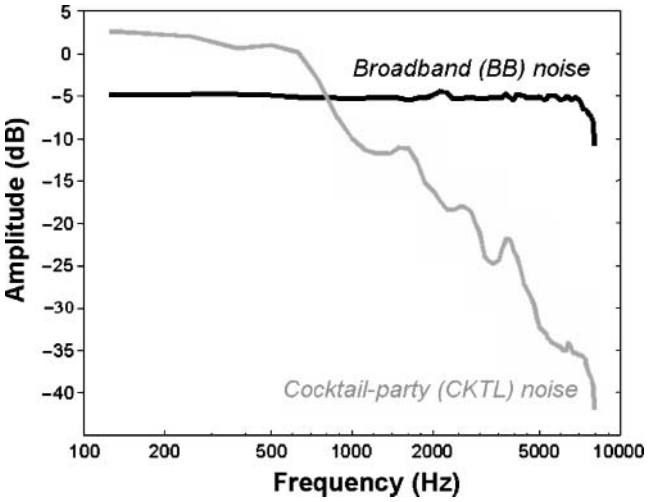
They participated in a game requiring the collaborative exchange of 16 target words with a speech partner seated 2 m in front of them (see Appendix A for more details about the game).

They played the game in a sound-treated booth, in nine different conditions: one quiet reference condition (40 dB SPL of background noise) and two types of noise played at four intensity levels (62, 70, 78, and 86 dB SPL). The two types of noise were (a) a broadband noise (BB), whose energy was attenuated above 10 kHz, and (b) a nonintelligible cocktail party noise (CKTL) made from eight mixed voices, with a spectral energy concentrated below 800 Hz (see Figure 1). Both noises were selected from the BD_Bruit database (Zeiliger, Serignat, Autesserre, & Meunier, 1994).

The experiment consisted of three successive recording sessions of these nine conditions, during which three sound-immersion methods were tested.

In the first session (IM1), noise was played into closed headphones (Sennheiser HD250 Linear II) characterized by a quasi-flat transfer function between 50 Hz and 40 kHz, with a slight attenuation of –3 dB SPL around 200 Hz and 3 kHz. The audio signal of the speech partner's voice was mixed with the noise via an analog mixer and played back into the speaker's headphones at a level compensating for the global attenuation induced by headphones.

Figure 1. Spectrum profiles of the broadband noise and cocktail party noise used in the experiments.



In the second session (IM2), noise and the partner's voice were similarly played into closed headphones. In addition, the audio signal of the speaker was played back into his or her headphones at a level compensating for the global attenuation induced by headphones.

In the third session (IM3), noise was played to the participants over two loudspeakers (Tannoy System 600) that were positioned 1.5 m from them in each lateral direction and at the level of their ears. Noise was then removed from the speech recordings using a dedicated noise-canceling method (Ternström et al., 2002) based on the estimation of the transmission channel and on time-domain subtraction of the estimated noise from the recordings. More details about the principle and performance of this method are presented in Appendix B. Speakers were asked (and monitored) to remain still during each 2- to 3-min noise condition but were allowed to move and relax in between while the transmission channel was estimated during a calibration phase preceding each noise condition. The headset microphone (Beyerdynamic Opus 54) was firmly attached to the head to avoid any movement of the face from the microphone. The map used for the interactive game was placed on a stand in such a way that the speakers could see both the map and their speech partner by moving their eyes only, but not their head, and writing on the map could be achieved with wrist movements only. In addition, the sound-treated booth, the cardioid directivity of the microphone, and its short distance from the speaker's lips (5 cm) improved the signal-to-noise ratio of the recordings and optimized the performance of the noise-canceling method.

A 30-min break in between each recording session allowed speakers to rest, to have a warm drink, and to limit their vocal fatigue. For the same reason, speakers were regularly invited to drink water during the experiment. For each session, the quiet condition and the eight noisy conditions were randomly selected to avoid habituation to noise level and to prevent speakers from basing their evaluation of the noise perturbation on the playlist order. The random selection of the playlist order was made before conducting the whole set of experiments, and it remained the same for each speaker.

Technical Details

Before beginning the experiment, we calibrated the noise output levels to measure the same intensity in dB SPL at the speaker's ears, regardless of whether the noise was played into headphones or over loudspeakers. To calibrate the level of noise played over the loudspeakers we used a ½-in. pressure microphone (B&K 4165) and an artificial head at the speaker's place in the booth. We used an artificial ear (B&K 4153) to calibrate the noise level into the headphones. In both cases, the microphone signal was sent to a preamplifier (B&K 2669) and an amplifier (B&K Nexus 2690). Likewise, the level of the partner's voice played into the speaker's headphones was calibrated so that its sound pressure level in the headphones, measured with the artificial ear, was equal to that measured with a pressure microphone and an artificial head at the speaker's place in the booth. The additional self-monitoring feedback in the second session was similarly calibrated so that the sound pressure level of the speaker's own voice measured in the headphones with an artificial ear was equal to that measured (externally) with a pressure microphone at the speaker's ear. We calibrated only the global level of these feedbacks but did not compensate for room effect or for the mouth-to-ear transfer function.

The same cardioid headset microphone (Beyerdynamic Opus 54), placed 5 cm away from and in front of the mouth, was used in the three sessions to record the audio speech signal, which was then preamplified (RME Octamic) and sampled at 44.1 kHz and 16 bits (RME ADI 8 Pro converter and RME DIGI 9652 HDSP sound card). Each speaker's SPL was calibrated prior to the test by measuring it on a sustained vowel with a digital sound level meter and by recording the audio signal of this reference production.

Measurements and Analysis

After each condition, speakers were asked to answer and evaluate on perceptual scales the four following questions (translated from French):

1. The way they had perceived their own voice (from 1 = *very audible* to 5 = *barely audible*)

2. The way they estimated they had been perceived by their speech partner (from 1 = *very audible* to 5 = *barely audible*)
3. The way they had perceived their partner's voice (from 1 = *very audible* to 5 = *barely audible*)
4. How uncomfortable they had felt when speaking (from 1 = *not uncomfortable at all* to 5 = *very uncomfortable*).

Utterances, target words, syllables, and segments of the target words were manually segmented with Praat software (Boersma & Weenink, 2005). We used MATLAB for the estimation of mean intensity, fundamental frequency (F0) of all the target words' syllables, and vowel duration produced by each speaker in each condition. We computed the centroid of the speech spectrum (0–6 kHz) for each sentence.

We conducted a two-way analysis of variance with repeated measures using SPSS. Factor 1, Immersion Method, had three levels: (a) headphones (IM1), (b) headphones with additional self-monitoring feedback (IM2), and (c) loudspeakers (IM3). Factor 2, Noise Level, had five levels: (a) quiet and (b) 62, (c) 70, (d) 78, and (e) 86 dB SPL. We tested the main effect of the Immersion Method factor first. If it was significant, we then examined, using Bonferroni adjustments, the specific contrasts between sessions IM1 and IM3 and between sessions IM1 and IM2, to determine two things: (a) whether the headphones paradigm induces a significant effect on speech parameters as compared with the loudspeakers paradigm (IM1–IM3) and (b) whether additional self-monitoring feedback in the headphones (IM1–IM2) affects speech production significantly, by enhancing or reducing the difference between the headphones and loudspeakers paradigms (IM1–IM3). The following notation has been adopted to report statistical significance of these different tests: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, and *ns* ($p > .05$).

Results

All the observations and statistical results for acoustic and perceptual parameters in BB and CKTL noises are summarized in Figures 2 and 3.

Difference Between the Headphones and Loudspeakers Paradigms (IM1–IM3)

As illustrated in the right column of Figure 2, speech production and the speaker's perception were significantly different when CKTL noise was played in headphones or over loudspeakers.

For acoustic parameters, the difference between the headphones and loudspeakers conditions can be considered constant for all ambient noise levels (i.e., lines in dark and light gray are parallel for these parameters, represented in the right column of Figure 2). Playing CKTL noise into headphones induced an offset effect on vocal intensity of +7.8 dB SPL ($p < .001$), compared with vocal intensity produced when the noise was played over loudspeakers. This is significant in comparison with the studied Lombard effect. (Speakers indeed increased their vocal intensity by 17.2 dB SPL when they adapted from quiet 86 dB SPL of CKTL noise in session IM3.) Similarly, F0 was 1.2 tones higher, on average ($p < .001$), when speakers were immerged into CKTL noise with headphones; vowels were 11 ms longer ($p = .033$); and the centroid of the speech spectrum was 189 Hz higher ($p = .002$).

For perceptual parameters, however, the difference between the headphones and loudspeakers conditions varied with level of CKTL noise (see right column of Figure 3): In the quiet condition, no significant difference was found in the speaker's perception between both paradigms. In the noise condition, however, perception was more degraded for headphones than for the loudspeakers paradigm, and this contrast tended to increase from 62 to 78 dB SPL of CKTL noise, then to decrease at high noise level (86 dB SPL).

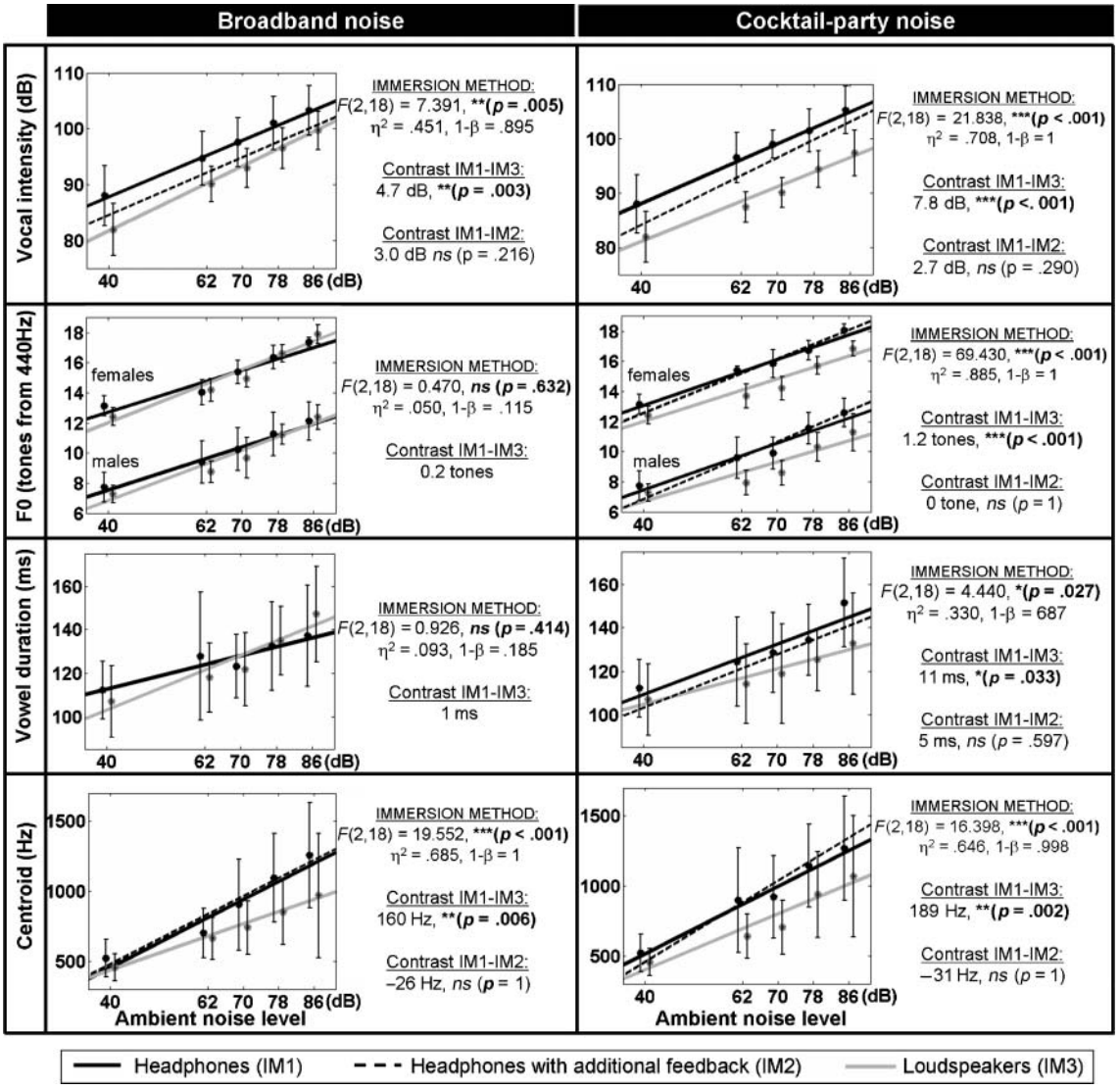
As illustrated in the left column of Figure 2, vocal intensity was significantly shifted by 4.7 dB ($p = .003$); similarly, the centroid of the speech spectrum shifted by 189 Hz ($p = .002$) when BB noise was played into headphones as compared with the loudspeakers paradigm (i.e., lines in dark and light gray are parallel). On the other hand, no significant difference was observed in BB noise between the loudspeakers and headphones paradigms for F0 and vowel duration, or for speaker's perceptual judgments (see right column of Figure 3).

Influence of Additional Self-Monitoring Feedback (IM1–IM2)

When additional self-monitoring feedback was returned in the headphones, vocal intensity tended to decrease by 3.0 dB SPL in BB noise ($p = .216$) and by 2.7 dB SPL in CKTL noise ($p = .290$; see dashed lines in the first row of Figure 2). There was also a very slight tendency for vowel duration to increase less in CKTL noise when additional feedback was returned in the headphones (5 ms, $p = .597$). Although these effects were not statistically significant and did not compensate for the difference between the headphones and loudspeakers paradigms, they still contributed to reduce that contrast.

On the other hand, additional self-monitoring feedback showed no effect on the other considered parameters. The centroid of speech spectrum was not modified (see dashed lines in the last row of Figure 2); neither were mean F0, the perception of one's own voice, the estimation of one's own intelligibility, the perception of the

Figure 2. Variation from quiet (40 dB SPL) to increasing levels of noise (62, 70, 78, and 86 dB SPL) of acoustic descriptors of speech. Three paradigms of immersion into noise are compared: (a) headphones (IM1, dark line), (b) headphones with additional self-monitoring feedback (IM2, dashed line), and (c) loudspeakers (IM3, light gray line). Results for the statistical comparison of these are indicated on the side of each graph. Mean values and interspeaker variability are represented for conditions IM1 and IM3. The lines represent the linear regression of data and summarize the main tendencies. Condition IM2 is always taken into account in the statistical analysis but is represented on the graph only when there was a significant difference between the headphones and loudspeakers paradigms (IM1 and IM3).



speech partner, and the evaluation of global discomfort in CKTL noise (see right columns of Figures 2 and 3).

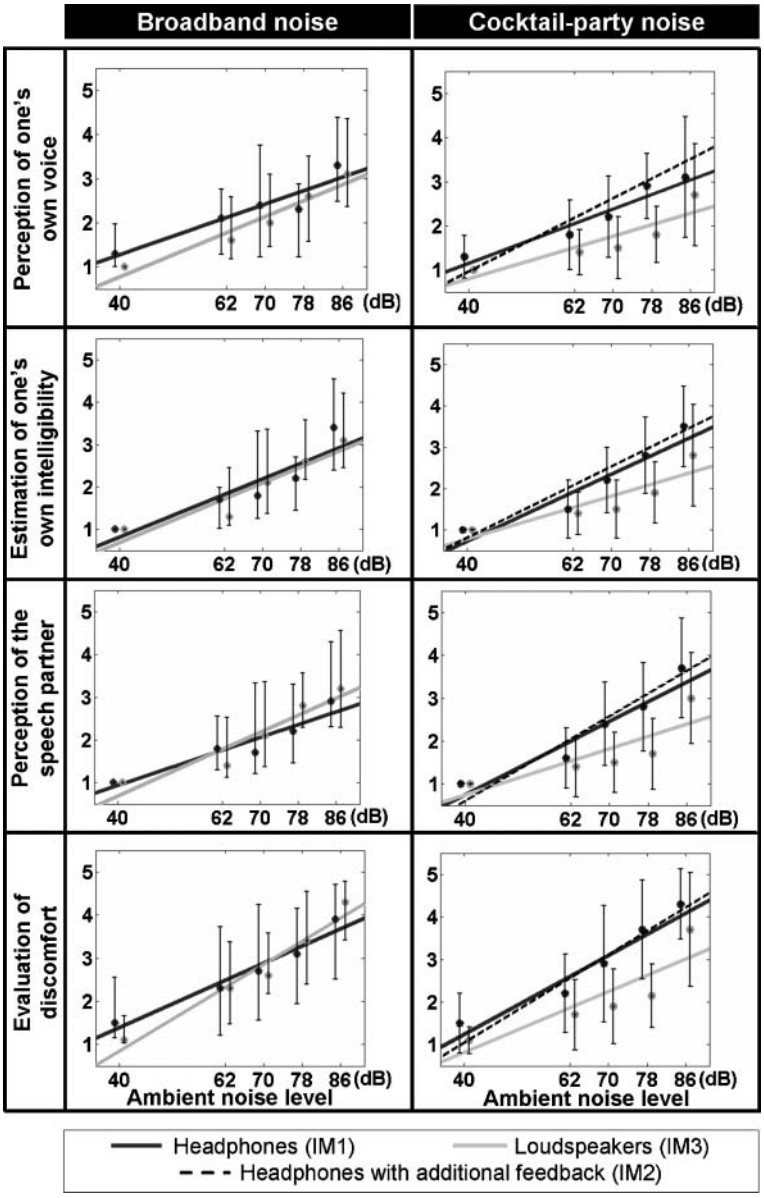
Discussion

Conclusions on the Influence of Sound Immersion Techniques

This experiment demonstrated that using closed headphones to immerse speakers in a noisy environment can have a significant impact on speech production and the speaker's perception, especially in CKTL noise. Overall, wearing headphones led to a globally more effortful phonation than did the loudspeakers paradigm. In CKTL noise it also degraded more considerably the speaker's perception and induced greater discomfort.

These results on acoustic and perceptual parameters are in accordance with results reported by Davis et al. (2006) showing that the amplification of articulatory movements is greater when noise is played over headphones than in a loudspeakers condition.

Figure 3. Variation from quiet (40 dB SPL) to increasing levels of noise (62, 70, 78, and 86 dB SPL) of perceptual evaluations by speakers. Three paradigms of immersion into noise are compared: (a) headphones (IM1, dark line), (b) headphones with additional self-monitoring feedback (IM2, dashed line), and (c) loudspeakers (IM3, light gray line). Mean values and interspeaker variability are represented for conditions IM1 and IM3. The lines represent the linear regression of data and summarize the main tendencies. Condition IM2 is represented on the graph only when there was a significant difference between the headphones and loudspeakers paradigms (IM1 and IM3).



In this experiment, we observed that additional feedback in the headphones tended to shift vocal intensity down, in accordance with previous results (Lane et al., 1961, 1970; Laukkanen et al., 2004). It had no or a negligible effect on other considered parameters, which is consistent with observations made by Kadiri (1998) on the first formant frequency but contrary to his observations on mean F0.

Impact on the Understanding of the Lombard Effect

Although the paradigm used to immerse speakers in noise (headphones vs. loudspeakers) showed a significant influence on speech production, it did not affect fundamentally the Lombard effect, that is, the modification of speech production from quiet to noise. Indeed, for the

acoustic parameters considered here, except for the centroid of the speech spectrum in BB noise, the difference between the headphones and loudspeakers paradigms remained similar in the quiet condition and in the different levels of ambient noise. This supports the idea that wearing closed headphones, even when no noise is played into them, induces an additional Lombard effect, by occluding the ear canal and attenuating the external auditory feedback. Further supporting this argument is the observation that wearing headphones induced modifications of voice and perception (greater vocal intensity, raised pitch, etc.) similar to those induced by noise exposure (i.e., Lombard effect).

These modifications, however, depended on the type of noise: The increase in vocal intensity between the headphones and loudspeakers conditions was, on average, 7.8 dB SPL for CKTL noise and 4.7 dB SPL for BB noise. Also, these modifications were smaller than the external-sound attenuation induced by the Sennheiser headphones used in that experiment, which measured approximately 10 dB SPL. The same tendency of a greater headphone effect in CKTL noise than in BB noise was observed for the other acoustical and perceptual parameters.

In addition, for CKTL noise the difference in speakers' perception between the headphones and loudspeakers paradigms appeared to be more complex than a simple offset, because this difference varied with noise level, increasing from quiet to 78 dB SPL of ambient noise and then decreasing at an extremely high level of noise (86 dB SPL).

Another argument in favor of a headphone influence being more complex than a simple attenuation of the external auditory feedback comes from the observation that additional self-monitoring feedback, which compensated for the global attenuation in intensity of the auditory feedback, had little influence on the difference between the headphones and loudspeakers paradigms and was not able to compensate fully for it.

These different observations may be explained by the fact that headphones not only attenuate the external auditory feedback but also modify the way noise is perceived by the speaker and how it masks their auditory feedback (resulting from both internal and external feedback). Indeed, the occlusion of the ear canal by headphones may improve bone conduction (Hood, 1962; Pörschmann, 2000; Von Bekezy, 1960) and enhance the low frequencies of the speakers' self-monitoring feedback. Because the energy of the CKTL noise is also concentrated in low frequencies, this could explain why wearing headphones had a greater influence on the speaker's perception and production in this kind of noise rather than in BB noise, where the energy is distributed over all frequencies. To explore this hypothesis further, it would be useful to replicate this experiment by applying a corrective filter on the self-monitoring feedback, returned in the speakers' headphones, to compensate for the degradation in high frequencies of the auditory feedback induced by the headphones.

Another possible explanation for these results is that the bias induced by the headphone paradigm is caused not only by the perturbed auditory feedback but also by the perturbed communication with a speech partner. Indeed, we observed in this first experiment that headphones affected not only the speaker's comfort and perception of his or her own voice but also that of the speech partner. Experiment 2 showed how communicative interaction with a speech partner influences the Lombard effect.

Experiment 2: Effect of Communicative Interaction

Material and Methods

Participants and Protcol

This experiment was conducted with 3 native French female speakers who ranged in age from 25 to 28 years. They were involved in a game that required the use of 17 fictional river names in a set carrying sentence ("La *rivière1* longe la *rivière2*"/"*The [river 1]* runs near *the [river 2]*"). More details about this game are presented in Appendix C.

Participants were asked to speak first in a quiet environment, then in the same cocktail party noise as used in the previous experiment. Noise was played at 85 dB SPL (calibrated at the speaker's ear) over two loudspeakers (A2t), placed 2 m apart from each other and both located 2 m away from the speaker.

For the three conditions, two sessions were recorded. In the first session, the speaker played the game alone and was asked to describe her actions aloud. In the second session, the speaker played the same game with a partner (the experimenter) placed 2.5 m in front of her. She had to give instructions to the partner, who in turn drew the instructed arrows on the board and asked for repetition when necessary. Speakers were not informed about the different conditions before the test; therefore, they were not aware during the first session (when they were alone) that they would have to interact with a speech partner in a later session.

Technical Details

The experiment was conducted in a sound-treated booth. The audio speech signal was recorded with a cardioid microphone (AKG C1000S) placed 50 cm away from and in front of the speaker's lips and digitized over 16 bits

at a sampling frequency of 44.1 kHz (Edirol M-100FX). The ambient noise was removed from the speech recordings using the same denoising algorithm (Ternström et al., 2002) as in Experiment 1 and as described in Appendix B. We estimated the impulse response of the transmission channel during a calibration step preceding every noise condition. Speakers remained seated and still during every noise condition. The sound-treated booth and the cardioid directivity of the microphone also improved the signal-to-noise ratio of the recordings and optimized the performance of the noise-canceling method.

We calibrated the vocal intensity level of each speaker before the test by measuring it on a sustained vowel with a digital sound level meter.

Lip movements were extracted from video recordings using a noninvasive labiometric method (Lallouache, 1990) that allows the individuals to speak very naturally and detects the whole lip contour with high precision.

Two 3CCD video cameras (JVC KY 15E) were placed in front and on the right side of the speaker, focusing on the speaker's lips, which were colored with blue lipstick. The speaker's head was completely immobilized by a helmet fixed to the wall. A vertical blue ruler was attached on the side of the helmet and served as a reference for the measurement of forward/backward movements of the lips. Cameras were synchronized, having a sampling rate of 25 images per second and a resolution of 638 × 582 pixels. Pictures were stored with betacam videotape recorders (Sony UVW 1400, UVW 1600, UVW 1800) and then digitized with a video capture card (Matrox Meteor I) at a rate of 50 video frames per second (with one picture corresponding to two interlaced video frames) and on 24 bits without data compression. These digitized pictures were processed with Traitement Automatique du Contour des Levres (TACLE) software, developed by Lallouache (1990). From the pictures, this program automatically detects the blue lips and the blue vertical ruler on the side using a chroma-key algorithm. It then applies classical pixel-based contour tracking algorithms to detect the internal and external lip contours. Several articulatory parameters were estimated from these contours, with a precision of 0.5 mm. To account for interspeaker variability in lip and face anatomy, we chose to normalize articulatory movements by the maximum articulatory gestures that each speaker was able to produce. With that aim, we asked speakers at the end of the experiment to open their mouth and then protrude their lips as much as possible. We then normalized all articulatory measurements to these extreme articulatory gestures, so articulatory data are presented in percentages instead of centimeters.

Measurements and Analysis

Utterances, target words, and their syllables were manually segmented from the audio signal using Praat software (Boersma & Weenink, 2005).

We estimated the mean intensity and F0 for all target words' syllables and their preceding article *la*. We had thus 102 measurements of these parameters for each condition and each speaker.

We estimated the centroid of the speech spectrum (0–6 kHz) over the 17 sentences produced by each speaker in each condition.

We measured the first formant frequency in the stable part of all the vowels [a] contained in the syllables [la] of target words and their preceding article, which gives 81 measurements for each condition and each speaker. For the same syllables [la] of the corpus, we also extracted the interlip area from the inner lip contour (see Figure 4) and measured the maximal amplitude of this articulatory parameter.

We measured maximal amplitude of lip compression on bilabial segments [m], [p], and [b], by calculating the difference between B′ (i.e., lip aperture from the external contour; see Figure 4) and $B'_{at\ rest}$ (i.e., when the lips are closed in a relaxed position, in between utterances). This articulatory parameter was defined only when the mouth is closed (i.e., when B, the lip aperture measured from inner contour [see Figure 4], is equal to zero). We had 16 measurements of this parameter for each condition and each speaker.

For the syllables [lu] and [la] of the corpus, we measured, respectively, the maximum and minimum amplitude of the protrusion movements of the upper lip. That articulatory movement was defined as the difference between P1 (most forward point of the upper lip; see Figure 4) and $P1_{at\ rest}$ (i.e., when the lips are closed in a relaxed

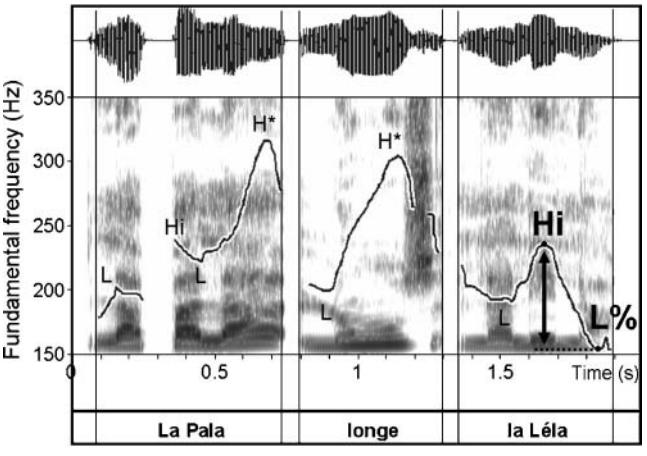
Figure 4. Articulatory parameters extracted from video recordings: interlip area (S) and protrusion of the upper lip (P1). Lip compression was defined as the difference between B′ and B, corresponding to lip aperture measured from external and inner contour of the lips.

position, in between utterances). For each speaker and each condition, we estimated the average contrast in protrusion between [u] and [a] vowels as the difference between the average (positive) protrusion of [u] vowels and the average (negative) protrusion of [a] vowels. Likewise, we estimated for each condition the average contrast between [a] and [u] vowels along the F1–F0 acoustic dimension, which is related to the perception of vowel height (Traunmüller, 1981).

For the 17 sentences produced by each speaker in each condition, we measured the intonation fall at the end of the utterance. Using Praat software, we detected the final low boundary tone of the utterance (L% of the model of French intonation proposed by Jun & Fougeron, 2000) and the preceding high tone (Hi of the secondary accent in that same model) and computed the interval between them in tones (see Figure 5).

Last, we examined the lengthening of the final syllable of the utterance. Research has established that syllable lengthening is a prosodic cue to discourse structure in French and marks the end of a prosodic unit (Delattre, 1966; Wenk & Wiolland, 1982). This lengthening is all the more important when this prosodic unit is of a high level (Bagou, Fougeron, & Fraunfelder, 2002; Christophe, Peperkamp, Pallier, Block, & Mehler, 2004); consequently, it is maximum on the final syllable of the utterance. The 17 target logatoms of our corpus were designed in such a way that all the syllables ([la], [le], [li], [ly], [lu], [lã], [pa], [ba], and [ma]) could be observed in both the final and penultimate positions of the utterances (e.g., "la Lalé longe la La<u>pa</u>"/"La Lila longe la <u>Pa</u>la"). For each speaker and each condition, this allowed us to measure how the same 17 syllables were lengthened when they were in the final position of the utterance, compared with when they were in the penultimate position.

Figure 5. Intonation fall at the end of utterances was measured as the difference between the last low (L) boundary tone (L% of the model of Jun & Fougeron, 2000) and the preceding high (H) tone (Hi of the secondary accent in that same model).



Results

Parameters Directly Related to Vocal Intensity

For the 3 speakers, vocal intensity and the other speech descriptors that can be related to it (F0, centroid of the speech spectrum, interlip area, F1) increased from quiet to noise in both interactive and noninteractive conditions (see Figure 6); however, this increase from quiet to noisy conditions always tended to be greater when speakers were interacting with a speech partner.

Other Parameters

Results are more complex for speech descriptors that are not directly related to vocal intensity but may instead be indicators of communicative strategies.

First, the 3 speakers increased lip compression on bilabial segments when communicating in noise with a speech partner. This can be considered a visible cue to bilabial place of articulation. Speakers enhanced the visible contrast in protrusion between [u] and [a] vowels as well as the audible contrast between them along the F1–F0 dimension. They also produced enhanced intonation falls at the end of utterances and lengthened more the final syllable of utterances compared with the quiet condition.

Observations in the noninteractive condition were more speaker and parameter dependent. Speaker 3 demonstrated behavior that we expected: She tended to increase most of these descriptors in noise for interactive condition only, yet she enhanced the lengthening of the final syllable in noise regardless of whether she interacted with a speech partner.

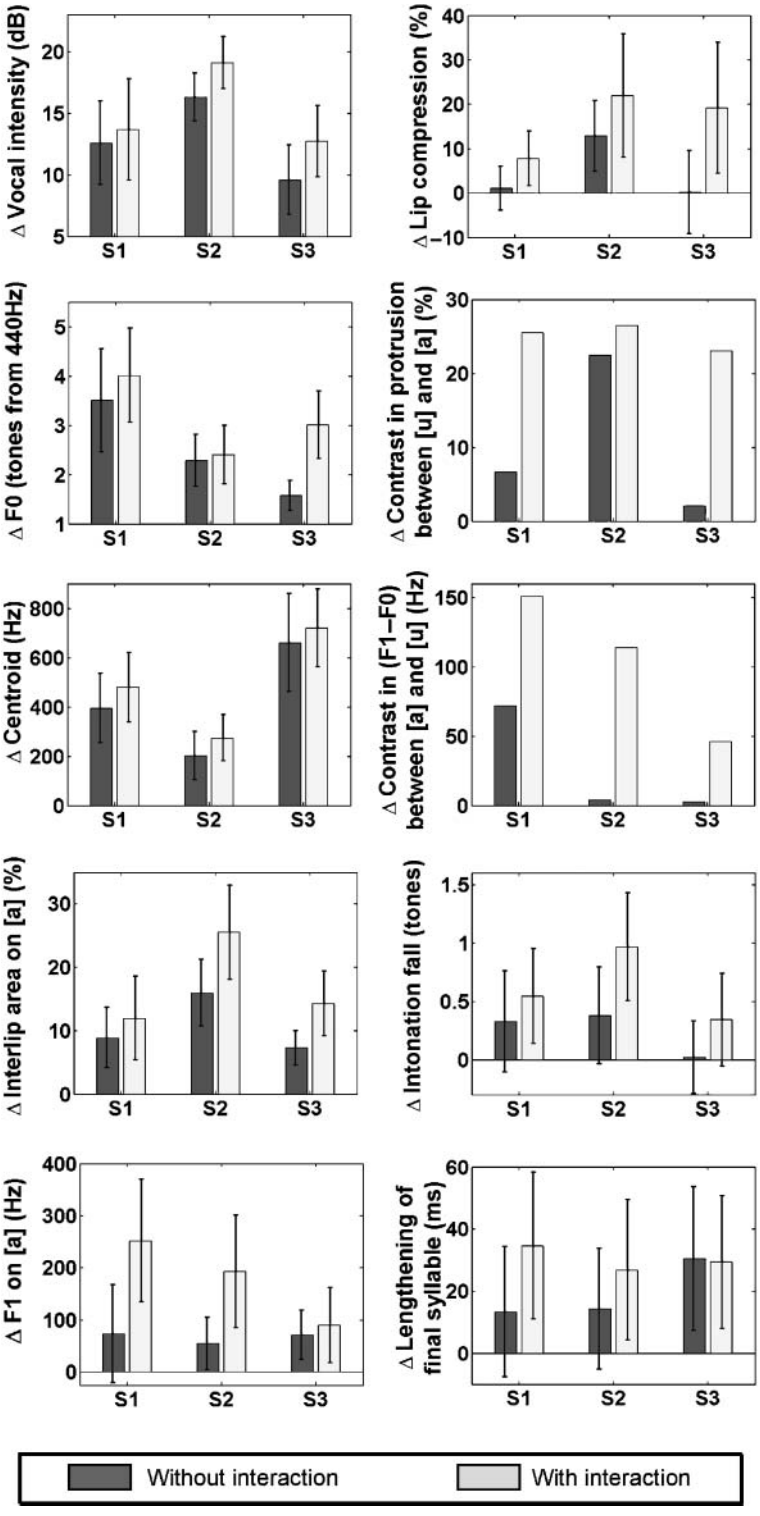
On the other hand, Speakers 1 and 2 tended to increase most of the descriptors in noise for both the interactive condition and the noninteractive conditions, although with a greater increase when they interacted with a speech partner. This behavior was similar to that observed for the parameters directly related to vocal intensity. However, Speaker 1 did not increase lip compression in noise in the case of noninteraction with a speech partner; neither did Speaker 2 enhance the acoustic contrast between [a] and [u] vowels.

Discussion

Conclusions on the Influence of Communicative Interaction

These results show that the effect of communicative interaction on speech production is more than simply an

Figure 6. Influence of communicative interaction on speech adaptation in noise for the three different speakers (S1, S2, and S3). Each bar represents the average variation of a speech descriptor from quiet to 85 dB SPL of cocktail party noise. Error bars stand for the intraspeaker variability in that adaptation over the different syllables or sentences, depending on the parameter considered. No error bar is indicated for the contrast between [a] and [u] vowels, because only one value of that contrast was estimated per condition and per speaker. Each graph compares whether the adaptation from quiet to noise is similar or different when speakers play the same game alone (dark bars) or in interaction with the experimenter (light gray bars).



offset—that is, an unvarying increase of speech parameters, regardless of whether speech is produced in quiet or in noise—but influences the studied phenomenon of speech adaptation from quiet to noisy conditions.

First, communicative interaction was found to influence the Lombard effect by amplifying the speech modifications that already occurred in noise, even without interaction. In particular, the increase of vocal intensity and related parameters was greater in interactive conditions, which is consistent with the results found by Amazi and Garber (1982) and Junqua et al. (1999). This also agrees with the greater increase of vocal intensity reported by studies of spontaneous or interactive speech compared with studies of read speech. Indeed, Korn (1954), Gardner (1966), and Kryter (1946) have computed the linear regression of vocal intensity as a function of ambient noise level and have reported slopes of, respectively, 0.30, 0.33, and 0.38 for spontaneous speech, whereas Lane et al. (1970) and Egan (1972) have measured slopes of, respectively, 0.12 and 0.15 for read speech. Similarly, speakers in Experiment 2 increased their vocal intensity by 15.2 dB SPL in average from quiet (40 dB SPL) to 85 dB SPL of noise when they interacted with a speech partner, which would correspond to a "slope" of 0.34 (= 15.2 ÷ 45), whereas they increased their vocal intensity by only 12.8 dB SPL in non interactive condition, corresponding to a slope of 0.28.

Second, communicative interaction was found to influence speech adaptation to noise by inducing additional modifications of speech that did not occur when the speaker did not interact with a speech partner. These modifications were speaker dependent, not directly related to vocal intensity, and may be considered as communicative strategies.

These different observations strongly support the idea of a communicative contribution to the speech adaptation in noise.

Impact on the Understanding of the Lombard Effect

Speech adaptation to noise was still observed in the absence of interaction with a speech partner. This concerned vocal intensity and related parameters, which agrees with the findings of most of the previous studies conducted on read speech (Castellanos et al., 1996; Garnier, Bailly, et al., 2006; Junqua, 1993; Stanton et al., 1988; Van Summers et al., 1988). Our results support the idea that speech adaptation in noise is neither a purely communicative effect nor a purely automatic regulation of voice intensity but instead a combination of both. This also concerned other speech descriptors that may rather account for language-specific strategies: enhanced contrast along the F1–F0 dimension for Speaker 1, increased lip compression for Speaker 2, enhanced contrast in protrusion and emphasized intonative fall for Speakers 1 and 2, lengthening of final syllables for the 3 speakers investigated. It appears that, even in unnatural laboratory conditions where speakers are not asked to address their speech to someone, they continue using strategies that they would normally use when interacting with another person. Consequently, in this second experiment we may have explored the influence of different levels of communicative involvement on the Lombard effect instead of comparing the Lombard effect in presence or absence of communicative motivation.

General Conclusion and Methodological Implications

Mechanisms Underlying the Lombard Effect

In this article we have presented two experiments in which we studied the influence of methodological paradigms and brought some new information to the understanding of the mechanisms underlying the Lombard effect.

In Experiment 1, we showed that speech adaptation from quiet to noise is influenced not only by the level and type of ambient noise but also by the method used for immersing the speaker into that noise, which may modify both the level and the spectrum of one's self-monitoring feedback. Because the influence of the immersion method also depended on the type of noise (greater in CKTL noise than in BB noise), speech adaptation to noise may be influenced by the way the auditory feedback is masked in frequency by the ambient noise, which supports the idea that the Lombard effect is partly a nonvoluntary regulation of vocal intensity influenced by the modification of the auditory feedback.

In Experiment 2, we showed that there can be speech adaptation to noise, even without any communicative interaction with a speech partner. However, an enhanced speech adaptation was observed from quiet to noise for an interactive condition. Furthermore, speech adaptation to noise consisted not only of acoustical and articulatory modifications that may be related to variations of vocal intensity but also of articulatory and prosodic modifications that could instead be interpreted in terms of communicative strategies to preserve intelligibility for the speech partner. These results not only argue in favor of a communicative contribution to the Lombard effect but also support the idea that the Lombard effect is not a purely automatic and uncontrollable regulation of voice,

or a purely communicative effect, but instead is a combination of both.

Methodological Implications

These results have some further methodological implications and may help in the choice of an appropriate paradigm to investigate Lombard speech, depending on the aims of the study, the type of noise used, and the observed parameters of speech production.

Choice of a Sound Immersion Technique

Experiment 1 illustrated that wearing headphones leads to a more effortful production but does not fundamentally modify the speech adaptation from quiet to noise. As a consequence, researchers using the Lombard effect as a natural way to force participants to increase their vocal effort, or researchers who are interested only in the global tendency of this effect, may not be concerned about the effect induced by headphones. On the other hand, this bias should not be neglected in studies that aim to make precise measurements of the Lombard effect for different levels of noise (Lane et al., 1970), to determine psychoacoustical mechanisms of this effect (Egan, 1972), or to characterize how a speaker reorganizes his or her communication strategies in different noisy situations (Garnier, 2007). In these last cases, the validity of using headphones still depends on the type of noise and the parameters considered: Indeed, we showed that wearing headphones in BB noise induces a negligible effect on F0 and vowel duration. Thus, the headphone paradigm can be chosen to explore the modification of these parameters in this noisy condition.

On the other hand, we showed that for most of the parameters, including vocal intensity in particular, using headphones to simulate a noisy environment in laboratory conditions can bias the adaptation from quiet to noise in a not-insignificant way, especially in CKTL noise. This effect of headphones on speech adaptation is consistent with the way headphones perturb the speakers' self-monitoring feedback and the perception of the partner. Thus, it can be interpreted as a perturbation of both audiological and communicative factors that precisely underpin the Lombard effect.

To compensate for this effect, a first solution could be to play additional feedback of the speaker's voice into headphones. We observed in Experiment 1 that such a technique reduced the influence of wearing headphones on speech adaptation in noise but was not able to compensate fully for it.

Another solution could be to model the effect of wearing headphones and to apply a corrective transformation to speech measurements. In a first approximation, which has been commonly made in previous studies (Dejonckere, 1979; Egan, 1972; Gardner, 1966; Korn, 1954; Kryter, 1946; Lane et al., 1970; Van Heusden, Plomp, & Pols, 1979), we can consider the variation of speech parameters as a linear function of the ambient noise level. This makes it possible to model the effect of headphones as a constant offset or as a linearly increasing function with ambient noise level (see Figure 2). However, this transformation might be different for varying models of headphones. Furthermore, this linear approximation of the Lombard effect is valid only on average, but no longer at an individual level (Garnier, Henrich, Dubois, & Polack, 2006). As a consequence, such a simple corrective transformation may be helpful in studies where speakers' adaptations to noise are averaged, but it would not be able to predict very precisely the speech of a peculiar speaker in natural conditions. Thus, in some cases it may be preferable to use loudspeakers instead of headphones as an experimental paradigm to immerse speakers into noise.

Choice of a Speech Production Task

Furthermore, we showed in Experiment 2 how communicative interaction influences the Lombard effect, not only by enhancing it but also by inducing additional modifications (increased lip compression, enhanced contrast between vowels along visible and audible dimensions, enhancement of prosodic cues to discourse structure). Researchers focusing on the audiological mechanisms underlying the Lombard effect may not be as interested in this communicative effect. On the other hand, studies that aim at characterizing speech produced in noise, in particular to improve the robustness of automatic speech recognition algorithms (Hansen, 1996; Junqua, 1993), or studies that deal with the increased intelligibility of Lombard speech (Junqua, 1993; Skowronski & Harris, 2006), should take into account communicative interaction in their experimental protocol; otherwise, they may miss some important phonetic characteristics and base their model on a type of speech different from the one encountered in a real situation by listeners or by automatic speech recognition systems.

Acknowledgments

We thank Lucie Bailly, Marion Dohen, Hélène Lœvenbruck, and Pauline Welby for their collaboration on a previous study that was used as an experimental basis for Experiment 2. We are also very grateful to Christophe Savariaux and Alain Arnal for their valuable help in Experiment 2. Last, we thank warmly the 10 speakers of Experiment 1 as well as the 3 speakers of Experiment 2, who kindly agreed to participate in this project, despite the discomfort of the noisy situations.

References

Amazi, D. K., & Garber, S. R. (1982). The Lombard sign as a function of age and task. *Journal of Speech and Hearing Research, 25,* 581–585.

Bagou, O., Fougeron, C., & Frauenfelder, U. H. (2002). Contribution of prosody to the segmentation and storage of "words" in the acquisition of a new mini-language. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2002* (pp. 159–162). Aix-en-Provence, France: Laboratoire Parole et Langage.

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America, 119,* 2363–2371.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.2.28) [Computer program]. Retrieved from http://www.praat.org.

Brown, G., Anderson, A., Yule, G., & Shillcock, R. (1983). *Teaching talk*. Cambridge, England: Cambridge University Press.

Burzynski, C. M., & Starr, C. D. (1985). Effects of feedback filtering on nasalization and self-perception of nasality. *Journal of Speech and Hearing Research, 28,* 585–588.

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

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language, 51,* 523–547.

Conture, E. G. (1974). Some effects of noise on the speaking behavior of stutterers. *Journal of Speech and Hearing Research, 17,* 714–723.

Davis, C., Kim, J., Grauwinkel, K., & Mixdorff, H. (2006). Lombard speech: Auditory (A), visual (V) and AV effects. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of the Third International Conference on Speech Prosody* (pp. 248–252). Dresden, Germany: TUD Press.

Dejonckere, P. (1979). L'effet Lombard-Tarneaud objectivé [Objective Lombard-Tarneaud effect]. *Electrodiagnostic-Therapie, 16,* 87–95.

Delattre, P. (1966). Les dix intonations de base du francais [Ten basic intonation patterns of French]. *The French Review, 40*(1), 1–14.

Dreher, J. J., & O'Neill, J. (1958). Effects of ambient noise on speaker intelligibility for words and phrases. *The Laryngoscope, 68,* 539–548.

Egan, J. J. (1972). Psychoacoustics of the Lombard voice response. *Journal of Auditory Research, 12,* 318–324.

Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *The Journal of the Acoustical Society of America, 70,* 45–50.

Fairbanks, G. (1954). Systematic research in experimental phonetics: A theory of speech mechanism as a servosystem. *Journal of Speech and Hearing Disorders, 19,* 133–139.

Garber, S. F., & Martin, R. R. (1977). Effects of noise and increased vocal intensity on stuttering. *Journal of Speech and Hearing Research, 20,* 233–240.

Garber, S. R., Siegel, G. M., & Pick, H. L., Jr. (1981). Regulation of vocal intensity in the presence of feedback filtering and amplification. *Journal of Speech and Hearing Research, 24,* 104–108.

Gardner, M. B. (1966). Effect of noise system gain and assigned task on talking levels in loudspeaker communication. *The Journal of the Acoustical Society of America, 40,* 955–965.

Garnier, M. (2007). *Communiquer en environnement bruyant: De l'adaptation jusqu'au forcage vocal* [Communication in noisy environments: From adaptation to vocal loading] (Unpublished doctoral dissertation). Université Pierre et Marie Curie, Paris, France. Retrieved from http://tel.archives-ouvertes.fr/tel-00177691.

Garnier, M. (2008). May speech modifications in noise contribute to enhance audio-visible cues to segment perception? In R. Göcke, P. Lucey, & S. Lucey (Eds.), *Proceedings of AVSP '08, the International Conference on Audio-Visual Speech Processing* (pp. 95–100). ISCA Archive. Retrieved from http://www.isca-speech.org/archive/avsp08/av08_095.html.

Garnier, M., Bailly, L., Dohen, M., Welby, P., & Lœvenbruck, H. (2006). An acoustic and articulatory study of Lombard speech: Global effects on the utterance. In *Proceedings of Interspeech '06, the International Conference on Spoken Language Processing* (pp. 17–22). Retrieved from http://www.isca-speech.org/archive/interspeech_2006/i06_1862.html.

Garnier, M., Dohen, M., Lœvenbruck, H., Welby, P., & Bailly, L. (2006). The Lombard effect: A physiological reflex or a controlled intelligibility enhancement? In H. C. Yehia, D. Demolin, & R. Laboissiere (Eds.), *Proceedings of ISSP '06, the 7th International Seminar on Speech Production* (pp. 255–262). Retrieved from http://hal.archives-ouvertes.fr/hal-00214307.

Garnier, M., Henrich, N., Dubois, D., & Polack, J. D. (2006). Est-il valide de considérer l'effet Lombard comme un phénomène linéaire en fonction du niveau de bruit? In *Proceedings of the 8th Congrès Français d'Acoustique* (pp. xxx–xxx).

Guastavino, C., Katz, B., Levitin, D., Polack, J. D., & Dubois, D. (2005). Ecological validity of soundscape reproduction. *Acta Acustica, 91,* 333–341.

Hansen, J. H. (1996). Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. *Speech Communication, 20,* 151–173.

Hanson, B. A., & Applebaum, T. H. (1990). Robust speaker-independent word recognition using static, dynamic and acceleration feature: Experiments with Lombard and noisy speech. In *Proceedings of ICASSP'90, the International Conference on Acoustics, Speech and Signal Processing* (pp. 857–860). doi:10.1109/ICASSP.1990.115973.

Hood, J. D. (1962). Bone conduction: A review of the present position with especial reference to the contributions of Dr. Georg von Bekesy. *The Journal of the Acoustical Society of America, 34,* 1325–1332.

Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology* (pp. 209–242). Dordrecht, The Netherlands: Kluwer Academic.

Junqua, J. (1993). The Lombard reflex and its role on human listener and automatic speech recognizers. *The Journal of the Acoustical Society of America, 93,* 510–524.

Junqua, J. C., Finckle, S., & Field, K. (1999). The Lombard effect: A reflex to better communicate with others in noise.

In *Proceedings of ICASSP '99, the International Conference on Acoustics, Speech and Signal Processing* (pp. 2083–2086). doi: 10.1109/ICASSP.1999.758343.

Kadiri, N. (1998). *Conséquences d'un environnement bruité sur la production de la parole* [Effect of noise exposure on speech production] (Unpublished doctoral dissertation). Université Paul Sabatier, Toulouse, France.

Kim, S. (2005). Durational characteristics of Korean Lombard speech. In *Proceedings of Eurospeech '05, the 9th European Conference on Speech Communication and Technology* (pp. 2901–2904). ISCA Archive. Retrieved from http://www.isca-speech.org/archive/interspeech_2005/i05_2901.html.

Korn, T. S. (1954). Effect of psychological feedback on conversational noise reduction in rooms. *The Journal of the Acoustical Society of America, 26,* 793–794.

Kryter, K. D. (1946). Effect of ear protective devices on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America, 18,* 413–417.

Lallouache, M. T. (1990). Un poste "visage-parole": Acquisition et traitement de contours labiaux [A "face-speech" system: Acquisition and processing of lip contours]. In *Proceedings of the 18th Journées d'Etudes sur la Parole* (pp. 282–286).

Lane, H. L., Catania, A. C., & Stevens, S. S. (1961). Voice level: Autophonic scale, perceived loudness and effects of sidetone. *The Journal of the Acoustical Society of America, 33,* 160–167.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research, 14,* 677–709.

Lane, H. L., Tranel, B., & Sisson, C. (1970). Regulation of voice communication by sensory dynamics. *The Journal of the Acoustical Society of America, 47,* 618–624.

Laukkanen, A. M., Mickelson, N. P., Laitala, M., Syrja, T., Salo, A., & Sihvo, M. (2004). Effects of HearFones on speaking and singing voice quality. *Journal of Voice, 18,* 475–487.

Leydon, C., Bauer, J. J., & Larson, C. R. (2003). The role of auditory feedback in sustaining vocal vibrato. *The Journal of the Acoustical Society of America, 114,* 1575–1581.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Dordrecht, The Netherlands: Kluwer Academic.

Lindblom, B., Brownlee, S., Davis, B., & Moon, S. J. (1992). Speech transforms. *Speech Communication, 11,* 357–368.

Lombard, E. (1911). Le signe de l'élévation de la voix [The sign of voice raising]. *Annales des Maladies de l'Oreille et du Larynx, 37,* 101–119.

Mixdorff, H., Grauwinkel, K., & Vainio, M. (2006). Time-domain noise subtraction applied in the analysis of Lombard speech. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of the Third International Conference on Speech Prosody* (pp. 94–97). Dresden, Germany: TUD Press.

Mokbel, C. (1992). *Reconnaissance de la parole dans le bruit: Bruitage/débruitage* [Speech recognition in noise: Noise degradation vs. cancellation] (Unpublished doctoral dissertation). Ecole Nationale Supérieure des Télécommunications, Paris, France.

Nonaka, S., Takahashi, R., Enomoto, K., Katada, A., & Unno, T. (1997). Lombard reflex during PAG-induced vocalization in decerebrate cats. *Neuroscience Research, 29,* 283–289.

Patel, R., & Schell, K. W. (2008). The influence of linguistic content on the Lombard effect. *Journal of Speech, Language, and Hearing Research, 51,* 209–221.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28,* 96–103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29,* 434–446.

Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R., & Kearney, J. K. (1989). Inhibiting the Lombard effect. *The Journal of the Acoustical Society of America, 85,* 894–900.

Pittman, A. L., & Wiley, T. L. (2001). Recognition of speech produced in noise. *Journal of Speech, Language, and Hearing Research, 44,* 487–496.

Pörschmann, C. (2000). Influences of bone conduction and air conduction on the sound of one's own voice. *Acta Acustica, 86,* 1038–1045.

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America, 85,* 295–312.

Siegel, G. M., Pick, H. L., Jr., Olsen, M. G., & Sawin, L. (1976). Auditory feedback in the regulation of vocal intensity of preschool children. *Developmental Psychology, 12,* 255–261.

Sinnott, J. M., Stebbins, W. C., & Moody, D. B. (1975). Regulation of voice amplitude by the monkey. *The Journal of the Acoustical Society of America, 58,* 412–414.

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

Södersten, M., Ternstrom, S., & Bohman, M. (2005). Loud speech in realistic environmental noise: Phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers. *Journal of Voice, 19,* 29–46.

Stanton, B. J., Jamieson, L. H., & Allen, G. D. (1988). Acoustic-phonetic analysis of loud and Lombard speech in simulated cockpit conditions. In *Proceedings of ICASSP '88, the International Conference on Acoustics, Speech and Signal Processing* (pp. 331–333). doi:10.1109/ICASSP.1988.196583.

Sundberg, J., & Nordenberg, M. (2006). Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech. *The Journal of the Acoustical Society of America, 120,* 453–457.

Ternström, S., Bohman, M., & Södersten, M. (2003). Very loud speech over simulated environmental noise tends to have a spectral peak in the F1 region. *The Journal of the Acoustical Society of America, 113,* 2296.

Ternström, S., Bohman, M., & Södersten, M. (2006). Loud speech over noise: Some spectral attributes, with gender differences. *The Journal of the Acoustical Society of America, 119,* 1648–1665.

Ternström, S., Södersten, M., & Bohman, M. (2002). Cancellation of simulated environmental noise as a tool for measuring vocal performance during noise exposure. *Journal of Voice, 16,* 195–206.

Ternström, S., Sundberg, J., & Collden, A. (1988). Articulatory F0 perturbations and auditory feedback. *Journal of Speech and Hearing Research, 31,* 187–192.

Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America, 85,* 901–906.

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *The Journal of the Acoustical Society of America, 69,* 1465–1475.

Tufts, J., & Frank, T. (2003). Speech production in noise with and without wearing hearing protection. *The Journal of the Acoustical Society of America, 114,* 1069–1080.

Van Heusden, E., Plomp, R., & Pols, L. C. W. (1979). Effect of ambient noise on the vocal output and the preferred listening level of conversational speech. *Applied Acoustics, 12,* 31–43.

Van Summers, W., Pisoni, B., Bernacki, H., Pedlow, R., & Stokes, M. (1988). Effect of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America, 84,* 917–928.

Von Bekezy, G. (1960). *Experiments in hearing*. New York, NY: McGraw-Hill.

Welby, P. (2006). Intonational differences in Lombard speech: Looking beyond F0 range. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of the Third International Conference on Speech Prosody* (pp. 763–766). Dresden, Germany: TUD Press.

Wenk, G., & Wiolland, F. (1982). Is French really syllable-timed? *Journal of Phonetics, 10,* 177–193.

Zeiliger, J., Serignat, J. F., Autesserre, D., & Meunier, C. (1994). BD_Bruit, une base de données de parole de locuteurs soumis á du bruit. In *Proceedings of the 10th Journées d'Etude de la Parole* (pp. 287–290).

Received July 9, 2008

Revision received January 20, 2009

Accepted September 19, 2009

DOI: 10.1044/1092-4388(2009/08-0138)

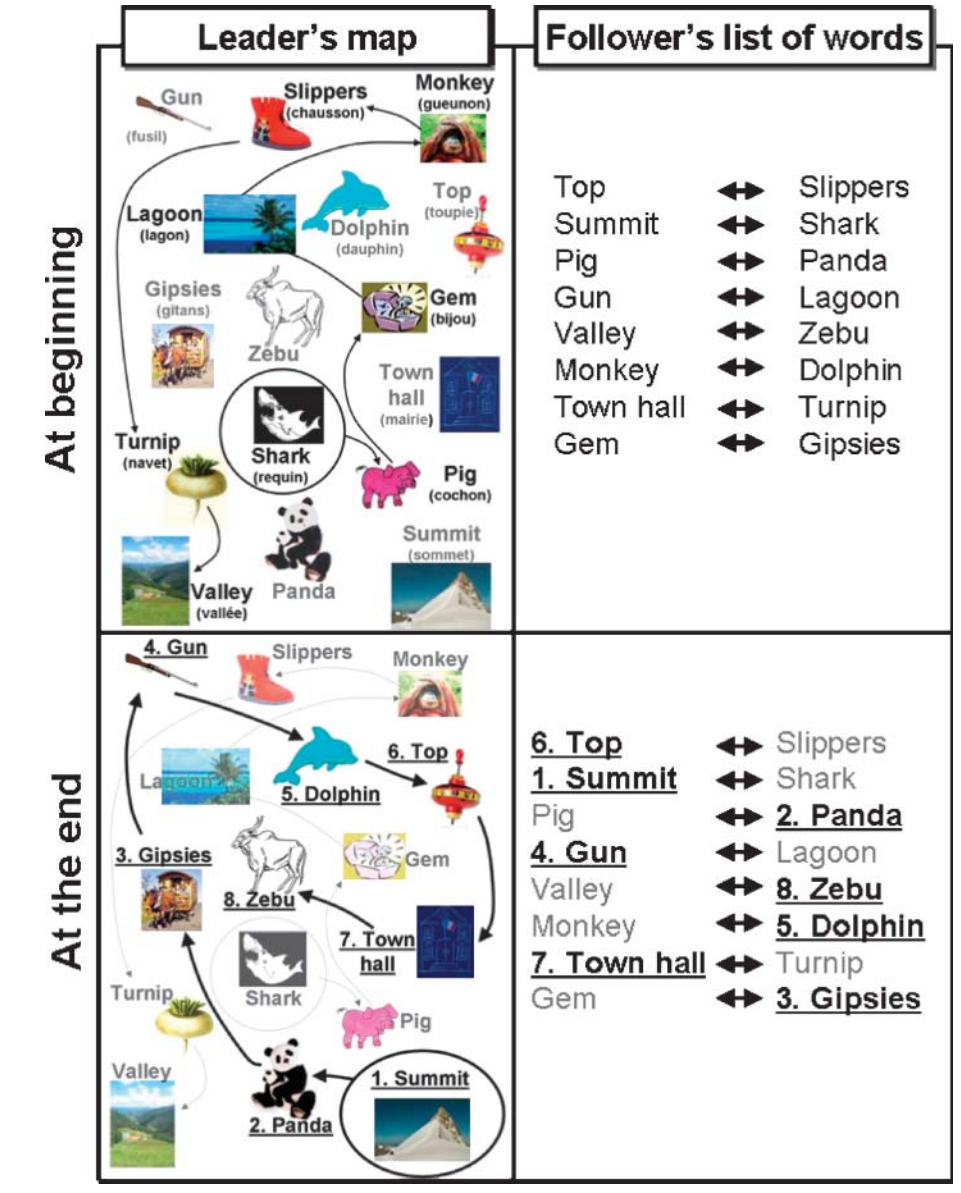
Contact author: Nathalie Henrich, GIPSA-Lab, Département Parole et Cognition, Grenoble, France. E-mail: Nathalie.Henrich@gipsa-lab.grenoble-inp.fr.

Appendix A (p. 1 of 2). First interactive game.

The game created for the Experiment 1 was inspired by Brown, Anderson, Yule, and Shillcock's (1983) Map Task game. It involves two participants: a leader and a follower.

The leader has a page on which 16 items are drawn and labeled, corresponding to the target words we want the speakers to produce. Before the game, 8 of these items are randomly selected by the experimenter and connected by arrows to form a path with a start point and an end point (see top left graph of Figure A1). The other 8 items are unconnected.

Figure A1. Example of the map and the corresponding list of pairs of words that both recorded speakers could have in the first interactive game, which was inspired by Brown et al.'s (1983) Map Task game. At the beginning of the game, a path is plotted on the leader's map (see top left graph). This path connects half of the drawn items, and the other items are left free. The pairs of words on the follower's list associate each connected item of the leader's map with an unconnected one (see top right graph). During the game, speakers have to exchange their complementary information to discover a new path linking the eight unconnected items of the leader's map (see bottom graphs).



Appendix A (p. 2 of 2). First interactive game.

The follower has a list of eight pairs of words, displayed in two columns. One word from each pair corresponds to a connected item on the leader's map, and the other word corresponds to one of the unconnected items (see top right graph of Figure A1). These pairs of words are randomly arranged; that is, words corresponding to connected items can be found in the left or in the right column, and words corresponding to the start and end points on the leader's map are not necessarily placed in the first or the last row. Thus, the follower cannot guess from the layout of the list which word the leader will say next.

To complete the game, the two participants have to exchange their complementary information to connect the eight remaining items and form a new path. To do this, the leader describes step by step to the follower the path that is already plotted on his map. For example, he could say "The first item is the shark." The follower then has to find the pair of words containing the word *shark* in his list and to answer the leader by telling him what the associated word is (*summit,* in this example). For instance, he could say "The shark is associated with the summit." Both partners then know that the start point of the new path is the "summit" item. Then, the leader tells the follower which item is in the second position, for example, by saying "Then I am going to the pig," to which the follower could answer "The pig goes with the panda." Both partners then know that the "panda" is the second item of the new path. Step by step, the leader will be able to draw the new path on his map (see bottom left graph of Figure A1), and the follower will be able to note this same path by numbering the items on his list (see bottom right graph of Figure A1).

After the leader has described completely the initial path and the follower has provided him with the corresponding items in his list, the new path is normally discovered by both partners. The leader has pronounced at least once the 8 target words corresponding to the initially connected items. The follower has pronounced all of the 16 target words. To conclude the game, the leader is then required to recapitulate the discovered new path by enumerating its items ordered from the start point to the end point and by ensuring that the follower agrees with him on the discovered path. Thus, the 8 remaining target words are pronounced by the leader.

During our experiments, the participants did not always immediately understand the task. For that reason, the protocol included playing a trial game with them before the recording session, and this was enough for them to understand fully the game's principle.

This game presents the advantage of simultaneously recording two speakers pronouncing a same set of target words in a semispontaneous way, with a real search for intelligibility to achieve a collaborative goal. On the other hand, this game does not allow exploration of some prosodic features, because neither the utterance structure nor the position of the target words within it are controlled.

Appendix B (p. 1 of 2). Denoising technique.

Principle

The noise cancellation method developed by Ternström et al. (2002) and implemented here consists of estimating the noise that would be recorded at the microphone if the speaker were quiet and then subtracting it from the noisy speech recording, sample by sample, in the time domain.

The estimated noise corresponds to the convolution of the noise signal played over the loudspeakers with the impulse response of the transmission channel, which includes the characteristics of the loudspeakers and the microphone as well as the room acoustics and the position of the person in the room.

The impulse response of the transmission channel can be estimated by playing a broadband noise over the loudspeakers and comparing it with its recording at the microphone. In our experiments, this characterization is made during a calibration step of 10 s preceding every noise condition.

Validity

The main problem of this method comes from the fact that its efficiency strongly depends on the accuracy of the channel characterization, and this may be affected by displacements of the person. To estimate the efficiency of the noise-canceling method and the extent to which movements of the speaker may affect acoustic measurements from the denoised audio signal, we conducted a short experiment.

We first recorded an utterance produced by a male speaker in a quiet sound-treated booth. This recorded sentence was then used to simulate a "virtual speaker" by means of a loudspeaker (TANOY Reveal) placed on a chair in the sound-treated booth. A cardioid headset microphone (Beyerdynamic Opus 54, also used in Experiment 1) was firmly attached to the virtual speaker so that the distance from the microphone remained constant.

The same sentence was played over the virtual speaker and recorded by the headset microphone in the following different conditions:

- In quiet, for reference.
- In noise, for the most ideal denoising conditions (condition T1), meaning the virtual speaker was motionless. The noise came from a different loudspeaker (Studer A1), placed 2 m in front of the virtual speaker. A cocktail party noise was first played at 85 dB, then a broadband noise was played at 79 dBC.
- In noise, for slight displacement of the speaker (condition T2). Instead of being simply placed on a chair, the virtual speaker was held by the experimenter seated on the chair. The experimenter demonstrated slight body movements similar to those occurring in Experiments 1 and 2.
- In noise, for deliberate forward–backward and left–right movements of the experimenter, seated on the chair and holding the virtual speaker (condition T3).

The characterization of the transmission channel was completed during a 10-s calibration step preceding each noise condition.

Several acoustic parameters were extracted from the T1, T2, and T3 recordings, before and after the denoising process. Methods used for these measurements were exactly the same as those used in the Experiments 1 and 2. We measured the mean intensity of every voiced and unvoiced segment of the utterance, the mean fundamental frequency of every syllable, and the first formant frequency of every vowel. We then calculated the average speech spectrum over the whole sentence (0–6 kHz) and measured its centroid. We compared these measurements with those from the reference recording, to evaluate the bias induced by noise or by denoising residuals. These comparisons are summarized in Table B1.

Estimation of F0 was barely affected by both types of noise. It could accurately be measured, even on the noisy signal. On the other hand, estimation of intensity was influenced by ambient noise, especially for unvoiced segments (1.6–9.0 dB). Similarly, the estimation of the vowels first formant and of the speech spectrum centroid were modified (respectively, from 4 to 26 Hz and from 54 to 179 Hz).

Denoising the recordings with the noise canceling method reduced very efficiently these biases for conditions T1 and T2, whereas the bias remained important for the condition T3. Thus, as long as we restrain the movements of the speaker, we can consider that acoustic measurements carried out from denoised recordings are not biased by more than 0.2 dB for intensity of voiced segments, 0.7 dB for unvoiced ones, 0.13 tones for F0, 10 Hz for the first formant frequency and 3 Hz for the centroid of the speech spectrum.

Appendix B (p. 2 of 2). Denoising technique.

Table B1. Evaluation of the bias induced by noise or denoising residuals on the estimation of several acoustic parameters, for different recording conditions: two types of noise (broadband noise and cocktail party noise) and different levels of speaker's displacement (no movement [T1], slight movements [T2], or large movements [T3]).

	Recording conditions					
	Broadband noise			Cocktail party noise		
Parameter	T1	T2	T3	T1	T2	T3
Δ SPL of voiced segments (dB)						
Before denoising	0 ± 0.03	0 ± 0.03	0.2 ± 0.1	0.1 ± 0.1	0.2 ± 0.1	0.9 ± 0.5
After denoising	0 ± 0.01	0 ± 0.01	0 ± 0.01	0.1 ± 0.1	0.1 ± 0.1	0.5 ± 0.3
Δ SPL of unvoiced segments (dB)						
Before denoising	1.6 ± 0.5	2.1 ± 0.6	3.6 ± 0.8	3.9 ± 1.8	3.5 ± 1.5	9.0 ± 2.0
After denoising	0 ± 0.03	0.1 ± 0.03	0.3 ± 0.1	0.2 ± 0.2	0.4 ± 0.3	6.6 ± 1.4
Δ F0 (tones)						
Before denoising	0 ± 0.05	0 ± 0.1	–0.1 ± 0.3	0 ± 0.2	0 ± 0.1	0 ± 0.04
After denoising	0 ± 0.02	0 ± 0.02	0 ± 0.02	0 ± 0.13	–0 ± 0.05	–0.1 ± 0.2
Δ Spectrum centroid (Hz)						
Before denoising	56	70	122	54	54	179
After denoising	3	3	6	3	3	122
Δ F1 (Hz)						
Before denoising	4 ± 4	5 ± 7	9 ± 9	10 ± 8	10 ± 8	26 ± 37
After denoising	2 ± 2	3 ± 2	4 ± 3	5 ± 4	5 ± 5	20 ± 26

Note. Reported values correspond to the difference between values measured from noisy recordings, before and after the denoising process, and values measured from the reference quiet recording.

Appendix C (p. 1 of 2). Second interactive game.

The game created for the second experiment is again derived from Brown et al.'s (1983) Map Task game. In this game, the recorded speaker always plays the role of leader. He gets a map on which 17 rivers are depicted, corresponding to the 17 target logatoms we want him to pronounce (see Figure C1).

To complete the task, the player has to connect these items with arrows, respecting the following three rules: (a) at the end of the game, no items must remain unconnected; (b) all items must have one single incoming arrow and one single outgoing arrow; and (c) the player has to always use the same carrying sentence—"La [*river1*] longe la [*river2*]" (The [*river1*] runs along the [*river2*])—to describe the arrow starting from [*river1*] and going to [*river2*].

These different rules allow the speaker to pronounce all the 17 target logatoms in the initial and final position of the utterance, corresponding to the subject and the object of a subject–verb–object utterance structure. For our purpose, the target words were logatoms derived from [lala] by varying one vowel or one consonant. In addition, the determiners la and the verb *longe* of the carrying sentence were chosen so that all vowels of the target words are in the same phonetic context [l_l] and all the consonants in the same context [a_a].

In the noninteractive experimental condition, the speaker draws the arrows on a map placed on a stand near him while describing aloud his actions. The experimenter monitors to ensure that rules are respected and all target words pronounced, but he does not show attention to the speaker and does not give him any feedback. In the interactive condition, the experimenter is standing a few meters in front of the speaker, facing a board on which the map is drawn. Instead of describing the arrows aloud, the speaker has to ask the experimenter to draw them for him on the board.

As in the first experiment, the protocol included playing a trial game with the speaker before the recording session so that all participants understand fully the game's principle.

This game presents the advantage of recording a set of target words pronounced still in a semispontaneous way but in a much more controlled phonetic context than with the previous game. It also allows the experimenter to vary the communicative conditions (alone or with interaction, varying the distance or the modalities of interaction between both speech partners, etc.). On the other hand, it is not possible with such a game to record more than one speaker at a time, which implies that the experimenter has to suffer the ambient noise for all the recorded participants and experimental conditions. In addition, the carrying sentence, which allows a greater control of prosody and segmental context reduces the spontaneity of the speaker's discourse.

Appendix C (p. 2 of 2). Second interactive game.

Figure C1. Example of the map that the speaker might have in the second interactive game, which also was inspired by Brown et al.'s (1983) Map Task game. He or she is asked to create a path connecting the different items, following some rules that are explained in detail in Appendix C.