Hyper-articulation in Lombard speech: An active communicative strategy to enhance visible speech cues?

Maëva Garnier,[1,a)] Lucie Ménard,[2] and Boris Alexandre[1,b)]
[1]*Centre National de la Recherche Scientifique, Laboratoire Grenoble Images Parole Signal Automatique, 11 rue des Mathématiques, Grenoble Campus, Boîte Postale 46, F-38402 Saint Martin d'Hères Cedex, France*
[2]*Département de Linguistique, Laboratoire de Phonétique, Center for Research on Brain, Language, and Music, Université du Québec à Montréal, 320, Ste-Catherine Est, Montréal, Quebec H2X 1L7, Canada*

(Received 22 August 2017; revised 12 July 2018; accepted 2 August 2018; published online 29 August 2018)

This study investigates the hypothesis that speakers make active use of the visual modality in production to improve their speech intelligibility in noisy conditions. Six native speakers of Canadian French produced speech in quiet conditions and in 85 dB of babble noise, in three situations: interacting face-to-face with the experimenter (AV), using the auditory modality only (AO), or reading aloud (NI, no interaction). The audio signal was recorded with the three-dimensional movements of their lips and tongue, using electromagnetic articulography. All the speakers reacted similarly to the presence vs absence of communicative interaction, showing significant speech modifications with noise exposure in both interactive and non-interactive conditions, not only for parameters directly related to voice intensity or for lip movements (very visible) but also for tongue movements (less visible); greater adaptation was observed in interactive conditions, though. However, speakers reacted differently to the availability or unavailability of visual information: only four speakers enhanced their visible articulatory movements more in the AV condition. These results support the idea that the Lombard effect is at least partly a listener-oriented adaptation. However, to clarify their speech in noisy conditions, only some speakers appear to make active use of the visual modality. © *2018 Acoustical Society of America*. https://doi.org/10.1121/1.5051321

[LK] Pages: 1059–1074

I. INTRODUCTION

It is now well known that seeing speech improves its perception (Dodd, 1977; Summerfield, 1992), especially when speech is perturbed in the acoustic domain, for example, for hearing-impaired individuals (Auer and Bernstein, 2007; Bernstein *et al.*, 2000; Conrad, 1977), for foreign listeners (Davis and Kim, 2001; Reisberg *et al.*, 1987), or when communicating in a noisy background (Bernstein *et al.*, 2004; Erber, 1975; MacLeod and Summerfield, 1987; Robert-Ribes *et al.*, 1998; Sumby and Pollack, 1954). However, how the visual modality is exploited in speech production, and whether speakers can make active[1] use of the visual channel (consciously or not) to improve their intelligibility, remain open questions. Noisy environments are typical situations in which speakers may adopt visual strategies. Such strategies not only make use of an alternative channel of transmission to compensate for the degraded intelligibility in the acoustic domain but could also enable speakers to limit their increase in vocal effort and prevent vocal damage. In this study, we explore whether speakers adapt the production of visible segmental cues to the available channels of interaction with a speech partner (auditory channel only vs both auditory and visual channels).

So far, many studies have shown that speakers adapt at least acoustically to noisy situations by talking louder and at a higher pitch (Castellanos *et al.*, 1996; Junqua, 1993; Van Summers *et al.*, 1988). Speech produced in noisy surroundings, also called "Lombard" speech, is characterized by higher first-formant (F1) frequencies of vowels, boosted energy above 2 kHz and increased vowel/consonant ratio in both vocal intensity and duration (Castellanos *et al.*, 1996; Garnier and Henrich, 2014; Junqua, 1993; Stanton *et al.*, 1988; Van Summers *et al.*, 1988). More recent studies have explored how these acoustic modifications affect segmental distinctiveness (Cooke and Lu, 2010; Garnier, 2008; Hazan *et al.*, 2012; Kim and Davis, 2014; Mixdorff *et al.*, 2007; Perkell *et al.*, 2007) and prosodic markers (Garnier *et al.*, 2006b; Patel and Schell, 2008; Welby, 2006).

All these acoustic modifications can be interpreted (1) as the consequences of automatic regulation of vocal intensity from the attenuated auditory feedback that a speaker gets from his/her own voice (Egan, 1972; Lombard, 1911; Tonkinson, 1994) and/or (2) as the expression of a listener-oriented adaptation that aims to maintain an acceptable level of speech intelligibility in the acoustic domain (Junqua *et al.*, 1999; Lane and Tranel, 1971).

Regardless of whether or not these acoustic modifications are produced actively in order to improve communication, perceptual tests have confirmed their positive influence on speech intelligibility in the auditory domain, for words and sentences (Chung *et al.*, 2005; Dreher and O'Neill, 1957; Lu and Cooke, 2008; Pittman and Wiley, 2001; Van

[a)]Also at: Université Grenoble-Alpes, F-38040 Grenoble, France. Electronic mail: maeva.garnier@gipsa-lab.grenoble-inp.fr
[b)]Also at: Laboratoire d'Étude des Mécanismes Cognitifs, Université Lumière Lyon 2, France.

Summers *et al.*, 1988), and more specifically, for vowels and voiced consonants (Junqua, 1993).

Fewer studies have shown that speakers also adapt to noisy situations by hyper-articulating speech: they not only increase the amplitude and speed of jaw, lip, and tongue movements (Alexanderson and Beskow, 2014; Fitzpatrick *et al.*, 2015; Garnier, 2008; Garnier *et al.*, 2006a; Kim *et al.*, 2005; Šimko *et al.*, 2016; Turner *et al.*, 2016) but also enhance articulatory contrasts between vowel categories (Garnier, 2008).

Articulatory movements are, of course, closely related to acoustic results (Lindblom and Sundberg, 1971). However, jaw and lip movements are very visible to a speech partner in a face-to-face interaction, whereas tongue, velum and larynx movements are not visible, or only partially and indirectly visible (Jiang *et al.*, 2002; Yehia *et al.*, 1998). Consequently, modifications of these barely visible movements can mainly be considered as the gestural bases for speech modifications in the acoustic domain, whereas amplified movements of the jaw and the lips may also reflect a listener-oriented strategy to improve speech intelligibility in the visual domain.

In any case, regardless of whether or not these visible cues are actively enhanced, perceptual tests have shown that they contribute to increased audiovisual (AV) intelligibility of Lombard speech compared to conversational speech produced in quiet conditions (Alexanderson and Beskow, 2014; Vatikiotis-Bateson *et al.*, 2007). In most cases, that increased AV intelligibility resulted from the benefit of a change from auditory-only (AO) to AV (i.e., AV-AO) in Lombard speech compared with conversational speech (three talkers of Davis *et al.*, 2006; Kim *et al.*, 2011; hard listening condition of Vatikiotis-Bateson *et al.*, 2007). In other cases, however, the AV intelligibility of Lombard speech was so high that the benefit (compared to AO) was smaller in Lombard speech than in conversational speech (one talker of Davis *et al.*, 2006; Alexanderson and Beskow, 2014; easy listening condition of Vatikiotis-Bateson *et al.*, 2007).

These findings raise the question of whether the hyper-articulated speech observed in noisy environments is:

- (Hyp1) simply related to the automatic regulation of vocal intensity due to masked auditory feedback or
- the expression of a listener-oriented adaptation that aims to

  (Hyp2) improve speech intelligibility in the auditory domain only (with fortunate but involuntary consequences in the visual domain) or

  (Hyp3) improve speech intelligibility in both the auditory and visual domains.

In support of Hyp1, many studies have observed a Lombard effect [i.e., vocal adaptation to noise exposure (NE)] in young children (Siegel *et al.*, 1976) and animals (Manabe *et al.*, 1998; Sinnott *et al.*, 1975), as well as in non-interactive (NI) communication situations (Egan, 1972; Lombard, 1911). Even when they are aware of the phenomenon, adult speakers do not seem to be able to inhibit it entirely (Pick *et al.*, 1989). Furthermore, many characteristics of Lombard speech are similar to loud or shouted speech: in particular, greater sound pressure level and higher fundamental frequency, flatter spectral tilt and higher F1 frequencies (Bond and Moore, 1990; Huber *et al.*, 1999; Lienard and Di Benedetto, 1999; Rostolland, 1982a, 1982b), but also greater amplitude and speed of jaw and lip movements (Geumann, 2001; Huber and Chandrasekaran, 2006; Schulman, 1989; Tasko and McClean, 2004). Only a few studies found these acoustic modifications to improve certain phonological contrasts, such as the F1 of vowels (Garnier, 2008; Junqua, 1993), the F0 of stop consonants (Hazan *et al.*, 2012), within-category dispersion (Cooke and Lu, 2010) or global vowel distinctiveness (Mixdorff *et al.*, 2007). Most of the studies, however, did not observe significant variations in between-category dispersion (Cooke and Lu, 2010; Kim and Davis, 2014) or within-category dispersion (Kim and Davis, 2014). Some of them even reported some reduced phonological contrasts (e.g., in the F2 of vowels, the voice onset time of stop consonants, the spectral mean of fricatives) and a more compact vowel space in Lombard speech (Bond *et al.*, 1989; Garnier, 2008; Hazan *et al.*, 2012; Perkell *et al.*, 2007). Furthermore, the type of noise was found to influence this vocal adaptation, but that influence was mainly quantitative, rather than qualitative. And the extent of vocal adaptation appears to be primarily related to the perceived loudness of the background noise rather than to the degree of energetic masking that the background noise exerts on speech (Cooke and Lu, 2010; Garnier and Henrich, 2014; Lu and Cooke, 2009).

In support of Hyp2 and Hyp3, however, some studies have shown that, although vocal adaptation is observed in NI communication situations, it is significantly greater in interactive situations (Amazi and Garber, 1982; Cooke and Lu, 2010; Garnier *et al.*, 2010). Several studies also showed that, even though Lombard speech shares many features with loud or shouted speech, it is also characterized by additional speech modifications that may not be directly related to the increase in vocal intensity, such as amplified lip closure, spreading and protrusion movements (Garnier, 2008; Turner *et al.*, 2016), enhanced frequency and amplitude modulation (Bosker and Cooke, 2018; Garnier and Henrich, 2014), and some enhanced contrasts (at the segmental, prosodic or pragmatic level) (Arciuli *et al.*, 2014; Beňuš *et al.*, 2015; Garnier, 2008; Garnier *et al.*, 2006b; Hazan *et al.*, 2012; Patel and Schell, 2008; Vainio *et al.*, 2012; Welby, 2006). Although most studies did not show a significant improvement in global between-category dispersion in Lombard speech, they did agree that formants tend to be produced more consistently in noise (i.e., with reduced within-category dispersion) (Cooke and Lu, 2010; Kim and Davis, 2014), which may contribute to improved vowel categorization. Furthermore, although speakers' primary strategy when coping with different noise types seems to consist of modulating their vocal intensity, more subtle and secondary speech modifications that are specifically adapted to the noise type and that enhance acoustic contrasts between speech and background noise were still observed, in addition to and when compatible with the primary shouting strategy (Garnier and Henrich, 2014). Finally, some of the speech modifications observed in Lombard speech, particularly hyper-articulation, are also observed in other kinds of clear

speech that are not specifically loud (infant-directed speech, foreigner-directed speech, hyper-visual speech, etc.) (DePaulo and Coleman, 1986; Freed, 1981; Green *et al.*, 2010). In these situations, such modifications are interpreted as communicative strategies to improve speech intelligibility in both auditory and visual domains. Consequently, it is reasonable to assume that this may be the case for Lombard speech too.

Several observations support Hyp3, i.e. the existence of communicative strategies that speakers may adopt to deliberately improve their intelligibility in the visual domain. In a study of eight participants, Fitzpatrick *et al.* (2015) reported that speakers opened their lips more when communicating in noise with a speech partner who could see them than when the partner could only hear them. Perceptually, the AV benefit for vowel identification was significantly greater for stimuli produced in face-to-face interaction than in an auditory interaction only condition. Another study showed that speakers adapted to the conditions of interaction by reducing speech overlap with their speech partner when the visual modality was not available, and especially when this interaction occurred in a noisy environment (Aubanel *et al.*, 2012). In mouthing, where the auditory modality is not available, speakers were shown to enhance lip protrusion (LP) compared to when vocalizing the same syllables (Bicevskis *et al.*, 2016), whereas their tongue movements were rather attenuated or unaffected by the act of mouthing. Finally, when comparing contrastive focus and clear speech produced by blind and sighted people, Ménard *et al.* (2014) and Ménard *et al.* (2016) observed similar F0, intensity, and acoustic contrast between vowels for both categories of speakers but significantly different strategies at the articulatory level, with greater lip contrasts for sighted people and, on the contrary, greater tongue contrasts for blind people.

On the other hand, certain observations cast doubt on the idea that speakers, or at least some speakers, are able to adapt to the available modalities of interaction. Thus, although Aubanel *et al.* (2012) observed that speakers reduced the speech overlap with their speech partner when the visual modality was not available, they did not find any other significant adaptation in their speaking style at the acoustic level (F0, intensity; F1, speech rate). Furthermore, the single speaker in our preliminary study (Garnier *et al.*, 2012) and the 14 speakers in the sample of Hazan and Kim (2013) did not show a general and significant tendency toward a greater enhancement of their visible lip gestures with NE when their speech partner could see them. On the contrary, the single speaker in the study of Garnier *et al.* (2012) demonstrated amplified articulatory movements when his speech partner could only hear him. The individual data in the study of Hazan and Kim (2013) also suggest that the small and non-significant effects observed at the group level may actually hide inter-individual differences in the communicative strategies used by speakers to clarify speech.

This study is intended to expand on the answers to these questions and test Hyp1, Hyp2, and Hyp3, by examining the following questions:

Q1. Are speech modifications induced by NE—particularly hyper-articulation—greater in interactive conditions, and are some of them even observed only in interactive conditions?
Q2. Are these speech modifications related only to the increase in vocal effort, or are there other speech modifications that cannot be directly related to voice intensity?
Q3. Does hyper-articulation affect only visible movements or all articulatory gestures?
Q4. Are visible movements more enhanced in noisy situations when the speaker can be seen by a speech partner?
Q5. Do speakers increase their vocal effort less in noisy situations when both audible and visible information are available (AV), compared to when they can only be heard by a speech partner (AO)?
Q6. Are barely visible movements comparably enhanced in noisy situations when the speaker can be seen vs only heard by a speech partner?

To that end, precise articulatory measurements of both lip and tongue movements were made, using electromagnetic articulography, as subjects produced speech in quiet surroundings and in the presence of babble noise, in three speech production conditions: a non-interactive condition (NI), a face-to-face interaction condition (AV), and an auditory-only interaction condition (AO). The results are discussed in light of our initial questions and hypotheses.

II. MATERIALS AND METHODS

A. Participants

Six native speakers of Canadian French (labeled S1 to S6) participated in the study. All of them were males, aged 22 to 45 years old. None of them reported any hearing or visual impairment. None of them were experts or students in phonetics or psychology, so we considered them as naive participants. They were only informed that they would undertake a speech production experiment. All participants gave written, informed consent in accordance with the institutional review board at Université du Québec à Montréal (UQAM) that approved this research project.

B. Speech material

The corpus consisted of ten repetitions of seven logatomes (/pap/, /pip/, /pup/, /pɛp/, /map/, /tap/, /nap/) that were embedded in the carrier sentence *Le mot __ me plaît* ("I like the word __"). The production order of the 70 sentences was chosen freely by the participant. Thus, the experimenter could not predict the target word and the participant really needed to adjust his intelligibility level to be understood.

C. Experimental procedure

Participants were recorded while speaking in six conditions: in a quiet environment vs in cocktail-party noise, in three communicative situations:

– NI condition: The speaker read the sentences aloud.

– AV condition: The speaker addressed his sentences to the experimenter, who was standing at a writing board placed 2 m in front of him. The two partners were facing each other. The experimenter interacted with the participant by writing down the items on the writing board. For about 5% of the utterances, the experimenter asked the participant to repeat because she had not understood the utterance or because she pretended not to have understood. Only the first occurrence of the utterance was analyzed. The experimenter (author M.G., female), of the opposite gender and unknown to the male participants, was the same partner for each of them, to minimize convergence effects and variations in interaction dynamics.
– AO condition: The experimenter was standing in the same place as in the AV condition and interacted similarly with the participant, except that this time, she faced the writing board and turned her back on the participant.

The participants experienced these six conditions in the same order: they began with the reading task (NI), first in noise, then in quiet. Then they experienced the most natural face-to-face condition (AV) (again, first in noise and then in quiet). They ended with the less natural AO condition.

The recordings were made in a sound-treated booth at the phonetics laboratory at UQAM. In the noisy conditions, noise was played over loudspeakers (Yamaha MSP7 Studio), located 1.5 m from the seated participant, in each lateral direction and at the level of his ears. The noise level was calibrated at 85 dBC at the participant's ears, using a digital sonometer (Scosche SPL1000). The cocktail-party noise came from the BD_Bruit database (Zeiliger *et al.*, 1994) and was composed of unintelligible mixed voices.

D. Measurements

The audio signal was recorded with a microphone (Shure SM58) placed 10 cm away from the speaker's lips, then digitized at a rate of 16 kHz. Noise was removed from the acoustic signal using a noise-canceling method proposed by Ternström *et al.* (2002) and used in previous studies of Lombard speech (Garnier and Henrich, 2014; Garnier *et al.*, 2010; Sodersten *et al.*, 2005).

The 3D movements of the lips, jaw and tongue were recorded synchronously with the audio signal, using 3D electromagnetic articulography (Carstens AG 500), at a rate of 200 Hz. The experimental setup was similar to the one used by Ménard *et al.* (2016). Four reference coils were placed behind each ear (coils 1 and 2 in Fig. 1) and just above the two upper incisors (coils 3 and 4 in Fig. 1). One coil was placed just under the lower incisors in order to examine jaw movements (coil 5 in Fig. 1). Four other coils were placed on the external contour of the lips (coils 6, 7, 8, and 9 in Fig. 1). The last three coils were placed on the central line of the tongue, approximately 1.5 (coil 10 in Fig. 1), 2.5 (coil 11 in Fig. 1), and 3.5 cm (coil 12 in Fig. 1) away from the tip of the tongue. These three points were chosen to examine the movement of, respectively, the tip (T), the body (B) and the root (R) of the tongue.

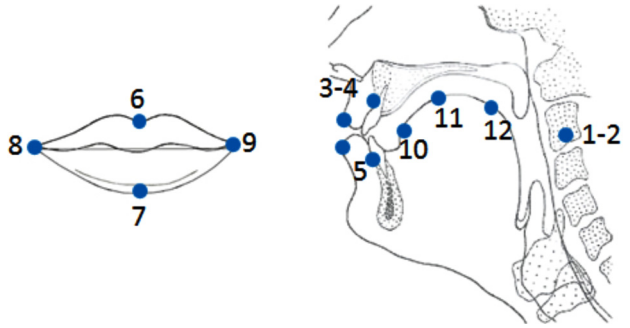

FIG. 1. (Color online) Position of the 12 coils used to track lip and tongue movements in 3D, using electromagnetic articulography (coils 1 and 2: reference on the left and right mastoids; coils 3 and 4: reference on the upper incisors; coil 5: lower incisors; coils 6 and 7: vermilion border of the upper and lower lips; coils 8 and 9: right and left lip corners; coils 10, 11, and 12: tongue).

E. Data post-processing

First, the 3D coordinates of the coils were transformed so they could be considered in the fixed reference frame of the participant's head, in order to interpret their displacement in terms of articulatory movements. The origin of this head frame was defined as the middle point between left and right ears (from the reference coils attached behind each ear: coils 1 and 2 in Fig. 1). The y-axis of this head frame was defined as the straight line passing through both ears. The x-y axial plane, corresponding to the fixed plane of the upper jaw, was defined from the two ear coils and from the middle point between the two reference coils glued above the upper incisors (coils 3 and 4 in Fig. 1). Consequently, these three points have a null z-coordinate.

A second coordinate transformation was then applied to the three coils attached to the tongue, in order to interpret their displacement relative to the lower jaw movements. These three coils were found to move almost in a plane. However, that plane did not correspond exactly to the x-z plane of the head frame. For each participant, we determined the actual sagittal plane of tongue displacement by conducting a principal component analysis of the displacement of the three tongue coils throughout the experiment (i.e., from at least 420 sentences). The 3D coordinates of these tongue coils were then converted into 2D by projection to this sagittal frame.

Using these 2D coordinates, we followed the JOANA method proposed by Henriques and van Lieshout (2013) to consider tongue movements relative to those of the lower jaw. This method relies on three reference coils: again, the two coils attached to the right and left ears (coils 1 and 2 in Fig. 1) and the coil attached to the middle of the lower incisors (coil 5 in Fig. 1), in order to estimate a frame of reference for the lower jaw and calculate, for each time step, the rotation angle between this lower jaw plane and the upper jaw plane, corresponding to the horizontal x-y plane of our previous reference head frame. Based on this time-varying angle, the 2D coordinates of the tongue movements in the actual sagittal frame could then be transformed, applying a rotation matrix, and expressed in the lower jaw frame.

F. Definition of the different time intervals

The beginning and end of each utterance were segmented manually from the audio signal, using PRAAT software. Acoustic and articulatory signals were then time-aligned and displayed for each utterance, using a graphic interface developed in MATLAB. This interface was used to manually segment the approximate time interval of the target word, from which the following elements were automatically detected:

- The maximum lip aperture (LA) on the vowel /o/, which always preceded the target word.
- The maximum LA on the central vowel of the target word (/a/, /i/, /u/, or /ɛ/).
- The minimum LA between these two maxima, corresponding to the initial consonant of the target word (/p/, /m/, /t/ or /n/).
- The minimum LA following the second maximum, corresponding to the final consonant of the target word (always /p/).

The syllable duration was defined as the time interval between the two LA minima preceding and following the target vowel.

G. Extraction of acoustic and articulatory descriptors

1. Acoustic descriptors

The mean intensity (SPL), mean fundamental frequency (F0), and mean frequency of the first two formants (F1, F2) were measured using PRAAT software, from the central 50 ms interval surrounding the voice intensity peak of each syllable (corresponding roughly to the time of maximum LA on the target vowel), using an autocorrelation method for F0 estimation and Burg's LPC algorithm for formant tracking. The standard deviations of F1 and F2 in each vowel category, /a/, /ɛ/, /i/, and /u/, were also calculated from the ten repetitions of each vowel by each speaker in each condition. The average variability observed within vowel categories will be referred to as the degree of "within-dispersion."

2. Descriptors of vowel articulation

LA was measured as the distance between the coils on the upper and lower lips (coils 6 and 7 in Fig. 1). Lip spreading (LS) was measured as the distance between the coils situated at both lip commissures (coils 8 and 9 in Fig. 1). That parameter could not be considered for speaker S3, since one of the commissure coils fell off during the experiment. The protrusion of the upper lip (LP) was measured as the distance between the coil on the upper lip (coil 6 in Fig. 1) and the origin of the head reference frame (i.e., the middle point between the ears: coils 1 and 2 in Fig. 1). The forward displacement of the tongue dorsum (TDx_jaw), was measured from the coil situated on the tongue dorsum (coil 11 in Fig. 1) and expressed in the time-varying sagittal frame of displacement of the lower jaw.

The maximum value of these articulatory descriptors was measured from the peak observed on each vowel. When no peak was observed (as was sometimes the case for tongue movements in particular), the value of the articulatory descriptor was measured at the time of maximum jaw aperture. Similar to the vowel formants, the standard deviations of LA, LS, LP, and TDx_jaw in each vowel category, /a/, /ɛ/, /i/ and /u/, were also calculated from the ten repetitions of each vowel by each speaker in each condition.

3. Descriptors of consonant articulation

Lip compression (LC) was measured on bilabial consonants as the minimum distance between the coils placed on the upper and lower lips (coils 6 and 7 in Fig. 1).

The forward displacement of the tongue tip (TTx) on apico-alveolar consonants was measured from the coil located on the tongue tip (coil 10 in Fig. 1), and expressed in the sagittal frame of the upper jaw. The maximum value of this articulatory descriptor was measured based on the peak observed on each consonant. When no peak was observed, the value of the articulatory descriptor was measured at the time of minimum jaw aperture.

Velocities of the lips and tongue tip at the occlusion release of bilabial consonants or apico-alveolar consonants (VL, VTT) were measured as the positive peak of the derivative signal of LA and from the derivative signal of the 3D displacement of the coil placed on the tongue tip (coil 10 in Fig. 1).

H. Statistical analysis

Several statistical analyses were conducted using R software. The conventional notation was adopted to report statistical results: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, and NS (not significant) $p > 0.05$.

An analysis of variance (ANOVA) was conducted for each acoustic and articulatory parameter in order to

- Examine the effect of NE (two levels: quiet and noise) on the value of these speech descriptors.
- Determine whether this speech adaptation to a noisy condition depended significantly on the condition of interaction (CI) [three levels: no interaction (NI), interaction in the audio domain only (AO), interaction in both audio and visual domains (AV)] and on the syllable (7 syllables considered for the global speech descriptors; 4 syllables /pap/, /pɛp/, /pip/, and /pup/ considered for the vowel descriptors; 5 syllables /pap/, /pɛp/, /pip/, /pup/, and /map/ considered for the descriptors of bilabial consonants; 2 syllables /tap/ and /nap/ considered for the descriptors of apico-alveolar consonants).

Since we expected participants to demonstrate different adaptation strategies to these experimental conditions, we first conducted individual statistical analyses for each participant. A group analysis was conducted as a second step, only when the majority of the speakers behaved similarly and when reporting a general tendency would be meaningful.

Individual analyses were conducted from a generalized model of the data (using the R package lm) comprising two fixed effects (of the NE and CI factors). Group analyses were done from a mixed model of the data (using the R package lme), taking into account not only fixed effects (of NE and CI) but also a random effect (of the speaker on the intercept).

For both individual and group analyses, we followed the same approach (favored by experts in statistics and explained by Bourne *et al.*, 2016): first, we searched for the simplest model to best explain the variance of a given parameter, using a descending approach (function step in R), based on minimization of the Bayesian information criterion. Hypotheses about the model's normality and homoscedasticity were validated by looking at the residual graphs. After examining the effects of the interaction terms remaining in the simplified model, we tested more specific contrasts, using the multcmp package in R and applying Bonferroni adjustments for multiple comparisons:

- the effect of NE in each condition: $NE_{/NI}$, $NE_{/AO}$, $NE_{/AV}$;
- the effect of communicative interaction on speech adaptation in noise: $NE_{/NI}$ vs $NE_{/(AO,\ AV)}$;
- the effect of the modality of interaction on speech adaptation in noise: $NE_{/AV}$ vs $NE_{/AO}$.

For the acoustic and articulatory descriptors of vowel production, we further tested:

- the inter-vowel contrast (in F1 and LA) between the close vowels /i/ and /u/ and the open vowel /a/, its variation with NE, communicative interaction and modality of interaction;
- the inter-vowel contrast (in F2, LS, LP, and TDx_jaw) between the back rounded vowel /u/ and the front unrounded vowel /i/, its variation with NE, communicative interaction and modality of interaction.

III. RESULTS

A. Effect of NE and communicative interaction

The results showed a general tendency toward increased voice intensity (SPL), F0 frequency, F1 frequency, LA, and syllable duration and a more forward position of the tongue dorsum (TDx_jaw) with NE (see Fig. 2). These modifications were significant for all the speakers in interactive conditions (on average: $\Delta SPL = +13.6$ dB ($z = 103.0$, $p < 0.001$) [Fig. 2(a)]; $\Delta F0 = +95$ Hz ($z = 66.1$, $p < 0.001$) [Fig. 2(b)]; $\Delta F1 = +188$ Hz ($z = 58.9$, $p < 0.001$) [Fig. 2(c)]; $\Delta LA = +8.4$ mm ($z = 59.4$, $p < 0.001$) [Fig. 2(d)]; $\Delta Duration = +32$ ms ($z = 1.8$, $p < 0.001$) [Fig. 2(e)]; $\Delta TDx\_jaw = +2.6$ mm ($z = 13.3$, $p < 0.001$) [Fig. 2(f)]). They were almost always smaller in the NI condition than in the interactive conditions (on average: -2.5 dB for SPL ($z = 11.1$, $p < 0.0001$) [Fig. 2(a)]; -55 Hz for F0 ($z = 21.9$, $p < 0.0001$) [Fig. 2(b)]; -67 Hz for F1 ($z = 12.1$, $p < 0.0001$) [Fig. 2(c)]; -4.3 mm for LA ($z = 17.6$, $p < 0.0001$) [Fig. 2(d)], -17 ms for syllable duration ($z = 5.5$, $p < 0.0001$) [Fig. 2(e)]; -1.7 mm for TDx_jaw ($z = 5.1$, $p < 0.001$) [Fig. 2(f)]). However, they still remained significantly non-null in the NI condition for almost all speakers (on average: $\Delta SPL = +11.1$ dB ($z = 59.4$, $p < 0.001$) [Fig. 2(a)]; $\Delta F0 = +40$ Hz ($z = 19.7$, $p < 0.001$) [Fig. 2(b)]; $\Delta F1 = +121$ Hz ($z = 26.8$, $p < 0.001$) [Fig. 2(c)]; $\Delta LA = +4.1$ mm ($z = 20.3$, $p < 0.001$) [Fig. 2(d)]; $\Delta Duration = +15$ ms ($z = 6.1$, $p < 0.001$) [Fig. 2(e)]; $\Delta TDx\_jaw = +0.9$ mm in NI ($z = 3.4$, $p = 0.003$) [Fig. 2(f)]).

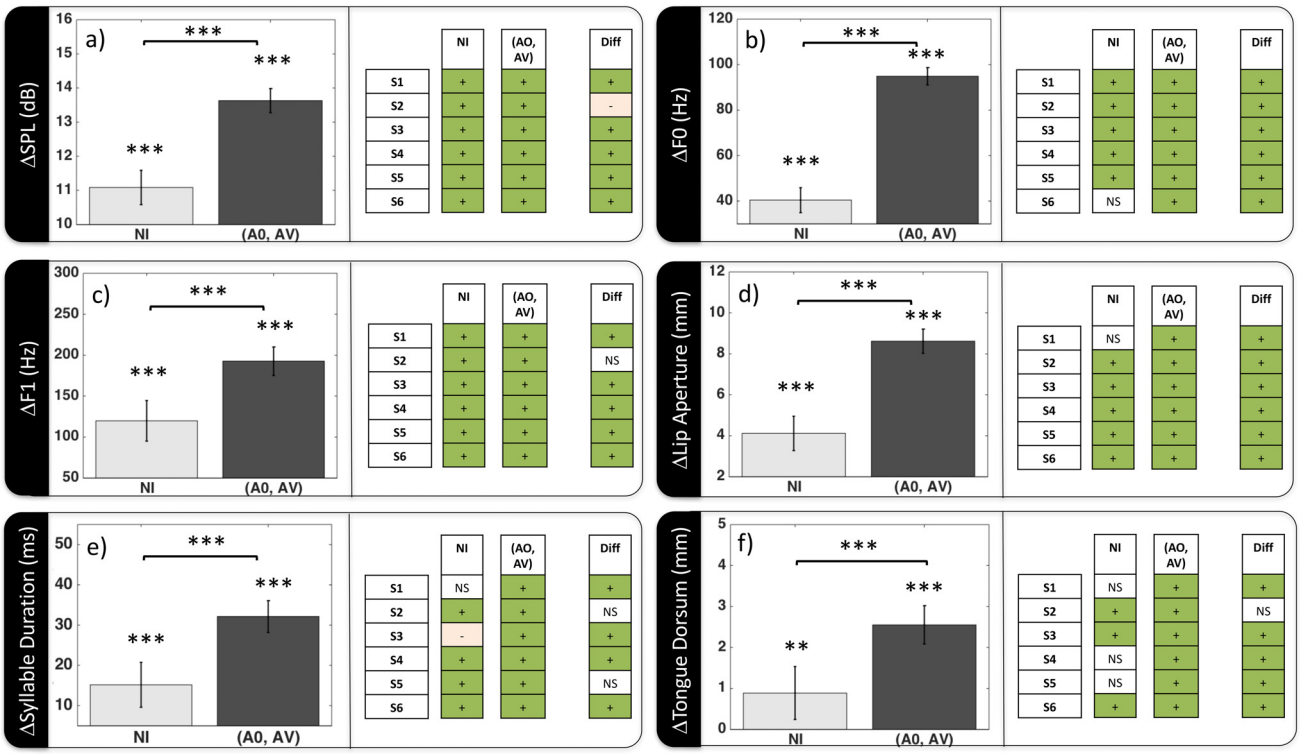

FIG. 2. (Color online) Variation with NE in (a) sound pressure level (SPL), (b) F0 frequency, (c) F1 frequency, (d) LA, (e) syllable duration, and (f) forward position of the tongue dorsum (TDx_jaw), observed in the NI condition and in the two interactive conditions (AO, AV) for all target words. The graphs represent the mean change observed for the whole speaker group. The error bars represent the confidence intervals estimated from the statistical model. The tables on the side of the graph summarize the variations observed for each participant (S1 to S6) in these conditions (+ for a significant increase, − for a significant decrease, NS for a non-significant change), as well as the difference in adaptation between interactive and NI conditions (diff).

Mean variations in F2, LS, and LP between conditions are presented in Fig. 3. A general tendency was observed toward an increase in F2 with NE for all vowels. The increase in F2 was systematic and very significant for the back rounded vowel [u] [on average: $\Delta F2 = +163$ Hz in NI ($z = 17.0$, $p < 0.001$); $+200$ Hz in AO and AV ($z = 23.9$, $p < 0.001$)] [Fig. 3(b)], whereas it was smaller for the other vowels ([a], [i], and [ɛ]) and significant only in some speakers [on average: $\Delta F2 = +37$ Hz in NI ($z = 5.2$, $p < 0.001$); $+74$ Hz in AO and AV ($z = 13.9$, $p < 0.001$)] [Fig. 3(a)]. In any case, these modifications were always similar or greater in interactive conditions, compared to the NI condition [on average: $+37$ Hz for F2 ($z = 4.8$, $p < 0.001$)].

On the other hand, no universal modification of LS and LP was observed with NE for all vowels, as for LA and tongue displacement. Nevertheless, a common tendency was observed across all the speakers toward increased LS with NE and decreased LP for the vowels [a], [i], and [ɛ] (on average: $\Delta LS = +0.8$ mm in NI ($z = 3.5$, $p = 0.002$) and $+1.8$ mm in AO and AV ($z = 11.8$, $p < 0.001$) [Fig. 3(c)]; $\Delta LP = -0.7$ mm in NI ($z = -4.7$, $p < 0.001$) and -1.2 mm in AO and AV ($z = 2.7$, $p < 0.001$) [Fig. 3(e)]), and decreased LS for the vowel [u] (on average: $\Delta LS = -0.9$ mm in NI ($z = -2.4$, $p = 0.06$) and -2.7 mm in AO and AV ($z = -10.5$, $p < 0.001$) [Fig. 3(d)]). Again, these different articulatory modifications were always similar or greater in interactive conditions than in the NI one (on average: $+1.0$ mm for LS on [a], [i], and [ɛ] ($z = 3.8$, $p = 0.0002$) [Fig. 3(c)] and -1.8 mm for LS on [u] ($z = -4.1$, $p < 0.001$) [Fig. 3(d)]; -0.5 mm for LP on [a], [i], and [ɛ] ($z = -2.2$, $p = 0.005$) [Fig. 3(e)]).

However, no common tendency was observed for the variation in LP on the vowel [u] with NE: some speakers significantly increased LP in noise (S2, S3, S4, S6), whereas others significantly decreased it (S1, S5) [see Fig. 3(f)]. For speakers S1, S2, and S4, these variations were significantly influenced by the communicative interaction and always resulted in a more accentuated protrusion of the vowel [u], or at least a less reduced one, in the interactive conditions than in the NI one (on average over the whole group: $+1.1$ mm, $z = 3.3$, $p = 0.002$) [Fig. 3(f)].

Figure 4 gives a general overview of the mean acoustic and articulatory modifications observed with NE for the four vowels [a], [ɛ], [i], and [u] in interactive and NI conditions, which enables the visualization of inter-vowel contrasts between open and close vowels, and between front and back rounded vowels. It shows that the vowel modification consists in a global shift of the vowel system toward significantly greater F1 values and slightly higher F2 values (especially for the back rounded vowel [u]), rather than a significant expansion of that system [Fig. 4(a)]. The acoustic contrast between open and close vowels tends to increase slightly in Lombard speech whereas the contrast between front and back vowels tends instead to decrease. The

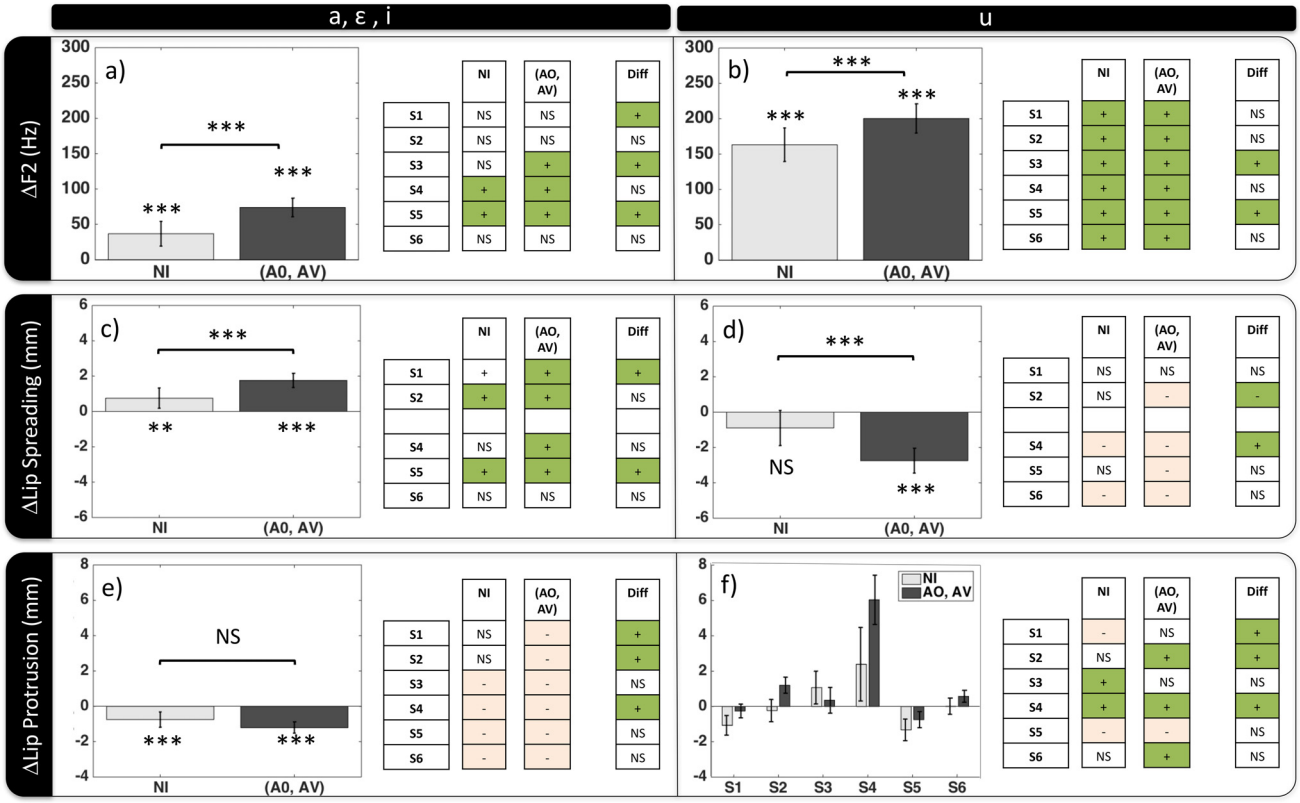

FIG. 3. (Color online) Variation with NE in (a), (b) second formant frequency (F2); (c), (d) LS; and (e), (f) LP observed in the NI condition and in the two interactive conditions (AO, AV), for the target words /pap/, /pip/, and /pɛp/ (left) and /pup/ (right). LS could not be measured for speaker S3, since one of the commissure coils fell off during the experiment. The graphs represent the mean change observed for the whole speaker group. The error bars represent the confidence intervals estimated from the statistical model. The tables on the side of the graph summarize the variations observed for each participant (S1 to S6) in these conditions (+ for a significant increase, − for a significant decrease, NS for a non-significant change), as well as the difference in adaptation between interactive and NI conditions (diff).

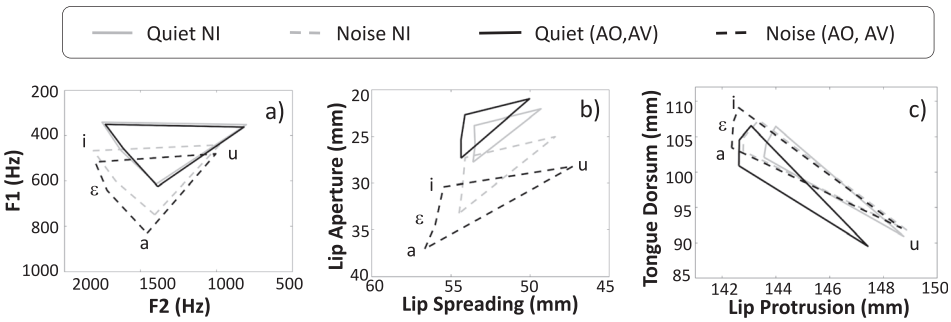

FIG. 4. Mean modification, over all six speakers, of the vowel system in the space of the first two formants (F1, F2), of LS vs LA, and of LP vs forward position of the tongue dorsum.

amplification of articulatory gestures in Lombard speech mainly concerns LA (for all vowels) and spreading (for spread vowels like [i] and [ɛ]) [Fig. 4(b)], rather than LP and TDx_jaw [Fig. 4(c)]. In agreement with the acoustic observations, the articulatory contrast between open and close vowels tends to increase slightly in Lombard speech. On the contrary, the front-back contrast, which tends to decrease somewhat for Lombard speech in the acoustic domain, is preserved or even increased in the articulatory domain, especially along the spreading dimension.

Mean variations in inter-vowel contrasts in both acoustic and articulatory spaces are presented with more detail in Fig. 5. As this figure shows, a general tendency was observed in the interactive conditions toward an increased inter-vowel contrast in F1 and LA between open and close vowels ([a] vs [i, u]) with NE (on average: ΔF1 contrast $= +68$ Hz ($z = 8.3$, $p < 0.001$) [Fig. 5(a)]; ΔLA $= +2.7$ mm ($z = 6.4$, $p < 0.001$) [Fig. 5(b)]); a reduced contrast in F2 between front and back rounded vowels ([i] vs [u]) (on average: ΔF2 contrast $= -140$ Hz ($z = -10.3$, $p < 0.001$)) [Fig. 5(c)]; and an increased contrast in LS and protrusion between front and back rounded vowels ([i]–[u]) (on average: ΔLS $= +4.1$ mm ($z = 11.1$, $p < 0.001$) [Fig. 5(d)]; ΔLP $= +1.7$ mm ($z = 6.0$, $p < 0.001$) [Fig. 5(e)]).

This same general tendency was also observed in the NI condition for the vowel contrast in LA ($+1.4$ mm on average, $z = 2.4$, $p = 0.044$) [Fig. 5(b)], F2 (-100 Hz on average, $z = -5.3$, $p < 0.001$) [Fig. 5(c)] and LS ($+1.5$ mm on average, $z = 2.8$, $p = 0.013$). On the other hand, no general tendency was observed in the NI condition for the vowel contrasts in F1 and LP, which still tended to increase with NE for four speakers (S3, S4, S5, S6) but decreased in the

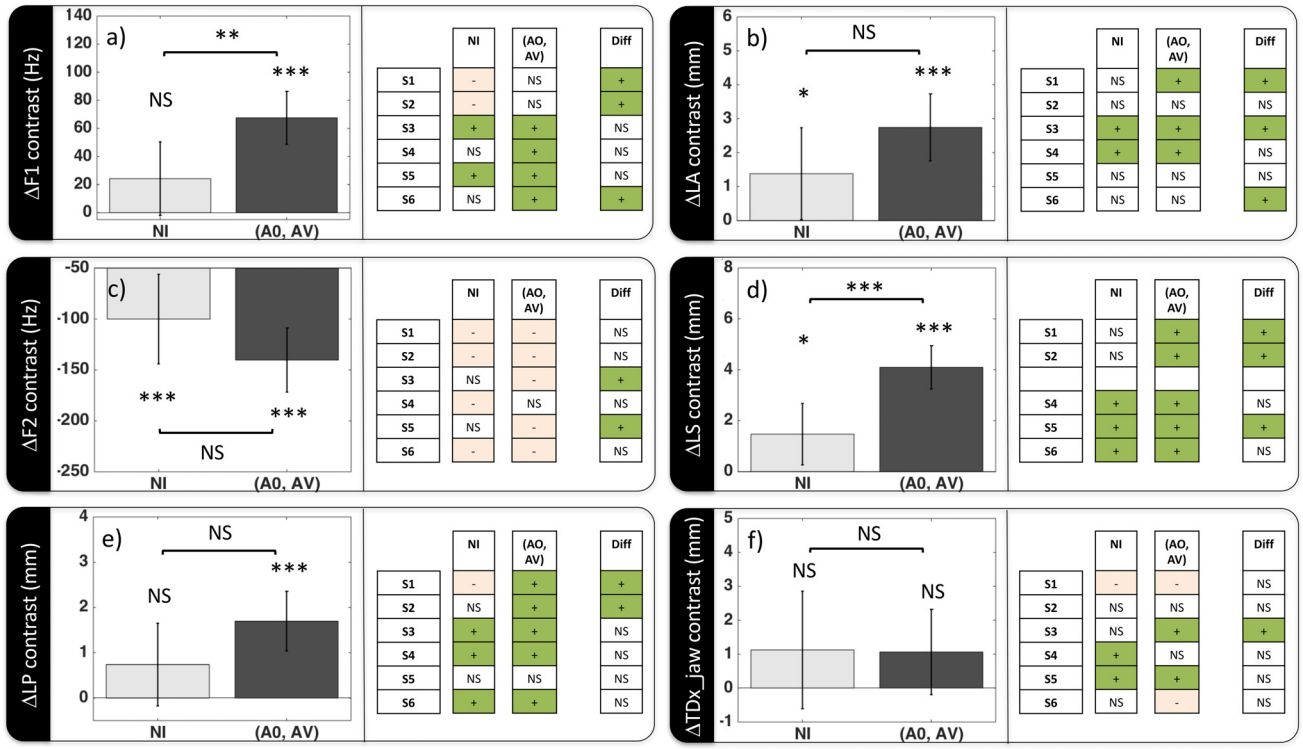

FIG. 5. (Color online) Variation with NE of the inter-vowel contrasts in (a) F1 and (b) LA, between the open vowel [a] and the close vowels [i] and [u], of the inter-vowel contrast in (c) F2, (d) LS, (e) LP, and (f) forward position of the tongue dorsum (TDx_jaw) between the front vowel [i] and the back rounded vowel [u]. The graphs compare this variation in the NI condition and in the two interactive conditions (AO, AV) for the whole speaker group. The error bars represent the confidence intervals estimated from the statistical model. The tables on the side of the graph summarize the variations observed for each participant (S1 to S6) in these conditions (+ for a significant increase, − for a significant decrease, NS for a non-significant change), as well as the difference in adaptation between interactive and NI conditions (diff).

absence of communicative interaction for the other two (S1, S2) [see Figs. 5(a) and 5(e)].

Finally, no general tendency was observed for the contrast in tongue dorsum position between front and back rounded vowels ([i] vs [u]): some speakers tended to increase that contrast with NE (S2, S3, S4, S5), whereas others tended to decrease it (S1, S6), without any reproducible effect of communicative interaction [see Fig. 5(f)].

Figure 6 presents the average variations with NE of the within-category dispersion for acoustic and articulatory parameters. The variability in F1 and F2 for the production of each vowel category increased slightly with NE (ΔF1 within-dispersion = +11 Hz on average, $z = 5.2$, $p < 0.0001$ [Fig. 6(a)]; ΔF2 within-dispersion = +15 Hz, $z = 5.2$, $p < 0.0001$ [Fig. 6(c)]), without any significant influence of the communicative interaction (for F1: 4 Hz, $z = 1.0$, $p = 0.32$ [Fig. 6(a)]; for F2: −6 Hz, $z = -1.0$, $p = 0.30$ [Fig. 6(c)]).

In agreement with these acoustic observations, the variability in LA, protrusion and TDx_jaw (tongue forward displacement) was also found to increase slightly with NE. For LA, it increased significantly in the interactive conditions only (ΔLA within-dispersion = +1.0 mm, $z = 6.9$, $p < 0.001$), but not in the NI condition (+0.3 mm, $z = 1.5$, $p = 0.26$), with a significant effect of the communicative interaction (+0.7 mm, $z = 2.8$, $p = 0.005$) [see Fig. 6(b)]. For LS, protrusion and TDx_jaw, the within-category variability increased significantly with NE (ΔLS within-dispersion = +0.6 mm on average, $z = 6.8$, $p < 0.001$ [Fig. 6(d)]; ΔLP within-dispersion = +0.2 mm, $z = 3.6$, $p < 0.001$ [Fig. 6(e)]; ΔTDx_jaw within-dispersion = +0.3 mm, $z = 2.0$, $p = 0$ 0.043 [Fig. 6(f)]), without any significant influence of the communicative interaction (for LS: 0.3 mm, $z = 1.8$, $p = 0.08$ [Fig. 6(e)]; for LP: 0 mm, $z = 0.2$, $p = 0.81$ [Fig. 6(e)]; for TDx_jaw: 0.2 mm, $z = 1.0$, $p = 0.31$ [Fig. 6(f)]).

Figure 7 presents the mean variations in consonant articulatory descriptors. A significant increase in LC (on the bilabial consonants [p] and [m]), more forward position of the tongue tip TTx (on the apico-alveolar consonants ([t] and [n]) and increased in lip and tongue tip velocities VL and VTT were systematically observed in the interactive conditions (on average: ΔLC = +1.1 mm ($z = 6.6$, $p < 0.001$) [Fig. 7(a)]; ΔVL = +117 mm/s ($z = 22.4$, $p < 0.001$) [Fig. 7(b)]; ΔTTx = +0.8 mm ($z = 2.7$, $p = 0.013$) [Fig. 7(c)]; ΔVTT = +26 mm/s ($z = 2.4$, $p = 0.029$) [Fig. 7(d)]), with a significantly greater amplitude compared to the NI condition for VL (on average: +83 mm/s, $z = 9.2$, $p < 0.001$) [Fig. 7(b)] but not for LC (on average: +0.3 mm, $z = 1.0$, $p = 0.57$) [Fig. 7(a)], TTx (on average: −0.4 mm, $z = -0.8$, $p = 0.42$) [Fig. 7(c)] or VTT (on average: +3 mm/s, $z = 0.2$, $p = 0.87$) [Fig. 7(d)]. Significantly non-null modifications were observed in the NI condition for at least half of the speakers (on average: ΔLC = +0.8 mm ($z = 3.4$, $p = 0.002$)

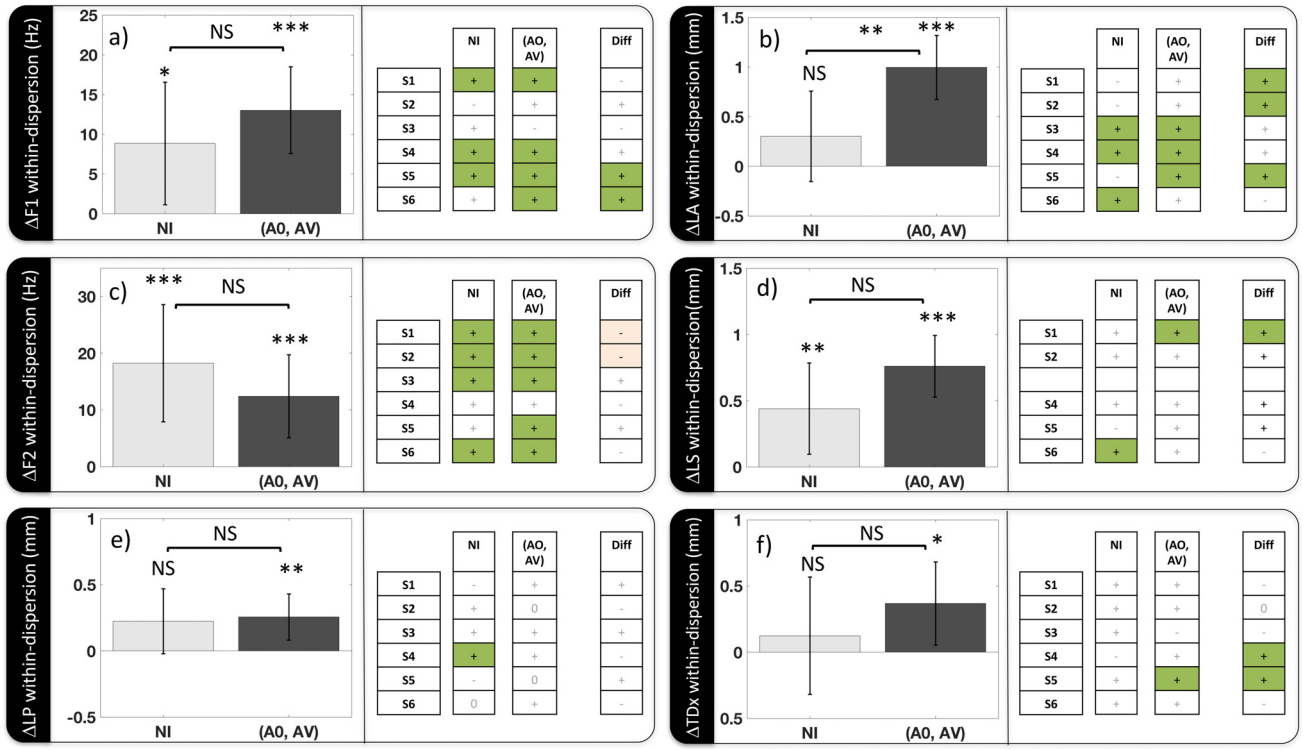

FIG. 6. (Color online) Variation with NE in the average within-category dispersion in (a) F1, (b) LA, (c) F2, (d) LS, (e) LP, and (f) forward position of the tongue dorsum (TDx_jaw). The graphs compare this variation in the NI condition and in the two interactive conditions (AO, AV) for the whole speaker group. The error bars represent the confidence intervals estimated from the statistical model. The tables on the side of the graph summarize the variations observed for each participant (S1 to S6) in these conditions (+ for an increase, − for a decrease), as well as the difference in adaptation between interactive and NI conditions (diff). Since statistical analyses could not be conducted on these parameters at the individual level, we arbitrarily chose to consider only frequency variations greater than 10 Hz and variations in movement amplitude greater than 1 mm; colored cells indicate a frequency variation greater than 10 Hz and a variation in movement amplitude greater than 1 mm.

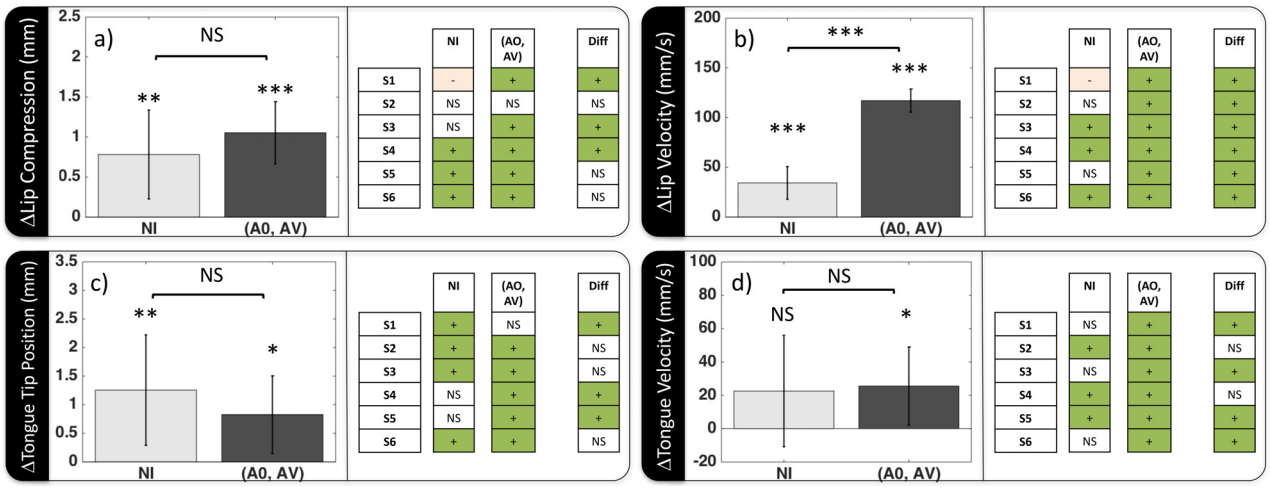

FIG. 7. (Color online) Variation with NE in (a) LC and (b) lip velocity on bilabial consonants /p/ and /m/, (c) forward position of the tongue tip (TTx), and (d) tongue velocity on apico-alveolar consonants /t/ and /n/. The graphs compare this variation in the NI condition and in the two interactive conditions (AO, AV) for the whole speaker group. The error bars represent the confidence intervals estimated from the statistical model. The tables on the side of the graph summarize the variations observed for each participant (S1 to S6) in these conditions (+ for a significant increase, − for a significant decrease, NS for a non-significant variation), as well as the difference in adaptation between interactive and NI conditions (diff).

[Fig. 7(a)]; $\Delta VL = +34$ mm/s ($z = 4.6$, $p < 0.001$) [Fig. 7(b)]; $\Delta TTx = +1.3$ mm ($z = 2.9$, $p = 0.007$) [Fig. 7(c)]; $\Delta VTT = +23$ mm/s ($z = 1.5$, $p = 0.24$) [Fig. 7(d)]).

B. Effect of the available sensory modality of interaction

As can be observed in Fig. 8 (right column), much smaller and subtler differences in speech adaptation to noise were observed between the two interactive conditions (AO and AV) than those between interactive and NI conditions. Furthermore, when significant differences were observed, they were speaker-specific: but no general tendency between the two interactive conditions could be highlighted at the group level. Although no general tendency was observed throughout the group, consistent variations in the different speech descriptors were observed for each speaker, supporting the idea that these variations are not random but reflect speaker-specific strategies.

Figure 8 summarizes these observations. To sum up, three groups of speakers could be distinguished:

- Speaker S1 (whose results were presented in a preliminary paper; Garnier *et al.*, 2012) differed from the others by demonstrating significantly greater speech modifications with NE in the AO condition than in the AV condition (+ symbols, highlighted in dark green in Fig. 8). The significant differences in speech adaptation were observed not only for global speech parameters that can be related to vocal effort (SPL, F0, F1, and LA) but also for other descriptors that may instead be related to vowel or consonant intelligibility (LA and TDx_jaw contrasts); these cues were visible or hardly visible.
- On the contrary, speakers S2, S3, S4, and S5 consistently demonstrated significantly greater global speech modifications (SPL, F0, F1, LA, LC, VL, LS, LP) with NE in the AV condition than in the AO condition (– symbols, highlighted in light orange in Fig. 8), and a smaller increase in the within-category dispersions (+ symbols) (also contributing to preserve intelligibility more in the AV condition). Only syllable duration was lengthened more for S4 and S5 in the AO condition. The speech parameters showing significantly different variations in the two interactive conditions varied from one speaker to another. However, in all cases, these significant differences concerned all kinds of speech descriptors, whether or not they were directly related to vocal effort, and whether they constituted easily visible or barely visible cues.
- Finally, speaker S6 demonstrated a kind of mixed behavior: he showed significantly greater speech modifications with NE in the AO condition for speech parameters that can be related to vocal effort (F0, F1 and articulatory velocities at occlusion release), but he enhanced protrusion gestures and contrast more in the AV condition.

IV. DISCUSSION

A. The Lombard effect: An automatic regulation of vocal intensity and/or a listener-oriented strategy to improve intelligibility?

In the introduction, we raised two initial questions:

Q1. Are speech modifications induced by NE—particularly hyper-articulation—greater in interactive conditions, and are some of them even observed only in interactive conditions?
Q2. Are these speech modifications related only to the increase in vocal effort, or are there other speech modifications that cannot be directly related to voice intensity?

In this study, speech modifications were indeed found to be significantly greater in the interactive conditions than in the NI condition for all, or almost all the speakers, as concerns the global voice parameters that can be directly related

			S1	S2	S3	S4	S5	S6	Group effect
Acoustic parameters (audible)	Global modifications	ΔSPL	+	-	NS	NS	-	NS	-0.4 dB z = -1.4, p = 0.17
		ΔF0	+	-	NS	NS	NS	+	+7 Hz z = 2.4, p = 0.018
		ΔF1	+	-	NS	NS	NS	+	+17 Hz z = 2.6, p = 0.009
		ΔF2 on [a], [i] and [ɛ]	NS	NS	NS	NS	NS	NS	+9 Hz z = 0.78, p = 0.68
		ΔF2 on [u]	NS	NS	NS	NS	NS	NS	+ 60 Hz z = 3.1, p = 0.004
	ΔInter-vowel contrasts	in F1 ([a] vs. [i],[u])	NS	NS	NS	NS	NS	NS	-34 Hz z = -2.1, p = 0.083
		in F2 ([i] vs. [u])	NS	NS	NS	NS	NS	NS	-28 Hz z = -1.0, p = 0.53
	ΔWithin-dispersion	of F1	+	-	+	+	-	-	0 Hz z = 0.07, p = 1.0
		of F2	-	+	+	0	+	+	+8 Hz z = 1.2, p = 0.44
ΔSyllable Duration			NS	NS	NS	+	+	NS	+3 ms z = 0.7, p = 0.46
Lip articulation (very visible)	Global modifications	ΔLip aperture (LA)	+	-	NS	-	-	NS	+0.1 mm z = 0.2, p = 0.84
		ΔLip compression (LC) on [p] and [m]	NS	NS	-	NS	NS	NS	-0.9 mm z = -2.8, p =0.013
		ΔLip velocity (VL)	+	-	-	NS	NS	+	5.8 mm/s z = 0.6, p = 0.83
		ΔLip spreading (LS) on [a], [i] and [ɛ]	+	NS		-	-	NS	+0.6 mm z = 2.0, p = 0.092
		ΔLip protrusion (LP) on [u]	NS	NS	NS	-	-	-	-1.7 mm z = -4.3, p < .0001
	ΔInter-vowel contrasts	in LA ([a] vs. [i],[u])	+	NS	NS	NS	NS	NS	+1.1 mm z = 1.3, p = 0.38
		in LS ([i] vs. [u])	NS	NS		NS	NS	NS	-0.1 mm z = -0.1, p = 0.98
		in LP ([i] vs. [u])	NS	NS	NS	NS	-	-	-1.6 mm z = -2.7, p = 0.015
	ΔWithin-dispersion	of LA	+	-	+	-	-	-	-0.1 mm z = -0.5, p = 0.88
		of LS	+	+		+	-	+	+0.4 mm z = 2.2, p = 0.072
		of LP	-	+	-	+	+	-	+0.3 mm z = 1.7, p = 0.20
Tongue articulation (hardly visible)	Global modifications	ΔTongue dorsum (TDx_jaw) on vowels	+	-	NS	NS	NS	NS	+0.7 mm z = 1.8, p = 0.15
		ΔTongue tip (TTx) on [t] and [n]	NS	NS	-	NS	NS	NS	-0.1 mm z = -1.1, p = 0.47
		ΔTongue velocity (VTT) on [t] and [n]	NS	NS	NS	NS	NS	+	+22 mm/s z = 1.1, p = 0.52
	ΔInter-vowel contrast	in TDx_jaw ([i] vs. [u])	+	NS	NS	NS	-	NS	-0.1 mm z = -1.4, p = 0.32
	ΔWithin-dispersion	of TDx_jaw	-	-	+	+	+	-	+0.6 mm z = 2.1, p = 0.092

FIG. 8. (Color online) Differences observed in the variation of speech parameters with NE between the AO and the AV interaction conditions, for the six speakers (S1–S6) in this study. Significantly positive differences (+), highlighted in dark green, indicate that greater adaptation was observed in the AO than the AV condition, whereas significantly negative differences (−), highlighted in light orange, indicate that greater adaptation was observed in the AV than the AO condition.

to voice intensity (SPL, F0, F1, LA, lip opening velocity). Despite this difference, these speech modifications remained significantly non-null in the NI condition for all, or almost all, the speakers.

All the speakers also demonstrated comparable or significantly greater speech modifications in the interactive conditions for speech parameters that are not directly related to voice intensity (e.g., syllable duration, F2, tongue displacement, LS

and LP, inter-vowel contrasts in F1 and LS). All these parameters but F2 did not consistently show a significant change with NE in the NI condition. However, all of them appear to be sensitive to the Lombard effect, even in the absence of communicative interaction, since they all showed a significant change with NE in the NI condition for at least two speakers.

These results therefore show that all six speakers were sensitive to the Lombard effect and significantly modified their speech when speaking in noisy surroundings, not only in interactive conditions, but also in the absence of communicative interaction. This confirms the findings of many previous studies (Castellanos *et al.*, 1996; Hazan *et al.*, 2012; Junqua *et al.*, 1999; Kim and Davis, 2014; Lu and Cooke, 2008; Šimko *et al.*, 2016; Van Summers *et al.*, 1988; Wassink *et al.*, 2007) and extends them to new and detailed observations of lip and tongue articulation.

The results also show that communicative interaction had a significant impact on speech adaptation to noise for all the speakers, leading them to modify their speech even more. This also confirms previous observations (Amazi and Garber, 1982; Cooke and Lu, 2010; Garnier *et al.*, 2010) and, again, extends them to lip and tongue articulation.

Furthermore, in both interactive and NI conditions, the speech modifications observed here consisted not only in an increase in voice parameters that can be directly related to the increase in voice intensity but also in an increase in some vowel- or consonant-specific descriptors and some inter-vowel contrasts that may not be directly related to voice intensity, in accordance with some previous studies (Cooke and Lu, 2010; Garnier, 2008; Junqua, 1993). This has two consequences: First, it supports the idea that the Lombard effect does not simply represent the automatic regulation of one's own voice intensity (Hyp1) but rather is a more complex adaptation that is exclusively or partially listener-oriented, aiming to improve speech intelligibility (Hyp2 and Hyp3) (Cooke *et al.*, 2014; Garnier *et al.*, 2010; Junqua *et al.*, 1999; Lane and Tranel, 1971). Second, it also rejects the idea, implicitly suggested in one of our previous papers (Garnier *et al.*, 2010), that these two contributions (automatic regulation of voice intensity vs listener-oriented enhancement of speech intelligibility), which correspond to different cognitive mechanisms, may actually be "distinguished" and that speech adaptation to a NI noisy condition may be underpinned by the first mechanism only, whereas speech adaptation to an interactive noisy condition may be underpinned by both mechanisms. Instead, it appears that both cognitive mechanisms always underlie speech production in noise and that speakers may still unconsciously try to improve their speech intelligibility when speaking in noise, even when they are not addressing a speech partner.

However, it is also important to mention that some speech modifications, such as a reduced contrast in F2 between front and back vowels or an increased within-category dispersion of acoustic and articulatory outcomes, were also observed with NE. Such modifications are not directly related to the increase in voice intensity but should contribute to decreasing vowel distinctiveness rather than improving it. Reduced phonological contrasts, in particular in F2, have already been reported in Lombard speech (Bond *et al.*, 1989; Garnier, 2008; Hazan *et al.*, 2012; Perkell *et al.*, 2007). However, previous studies observed no significant change in within-category dispersion (Kim and Davis, 2014) or a significant improvement (Cooke and Lu, 2010). The reduced contrast in F2 can be interpreted as the direct consequence of an increased LA, which dramatically affects the first two formants of back rounded vowels (Savariaux *et al.*, 1995). The increase in within-category dispersion may reflect decreased precision in speech motor control, due to the very attenuated auditory feedback that the speaker has of his own voice at very high noise levels, as in our experiment.

B. Do speakers make deliberate use of the visual modality to improve their speech intelligibility in noisy communication conditions?

A first element of an answer to that question is to determine whether hyper-articulation affects only visible movements or all articulatory gestures (Q3).

In this study, almost all articulatory gestures, whether visible (LA, LS for [a], [i], [ɛ]) or less visible (position of the tongue dorsum for vowels or of the tongue tip for apico-alveolar consonants, tongue tip velocity) remained similar or were enhanced by all the speakers when exposed to noise. Only LP for the vowel [u] and LC for bilabial consonants showed some inter-speaker variability: LP for [u] was enhanced by some speakers (S2, S3, S4, and S6) and reduced by others (S1 and S5). The amplitude of LC was enhanced by half of speakers (S4, S5, S6) and reduced by S1 in the NI condition. It did not change significantly for S2.

In any case, the visible contrasts in LA between open and close vowels and the visible contrast in LS and protrusion between spread and rounded vowels were conserved or enhanced in noisy surroundings by all speakers (except the contrast in LP for S1 in the NI condition). The less visible contrast in tongue dorsum position between front unrounded and back rounded vowels was also conserved or enhanced with exposure to noise but for four speakers only (S2, S3, S4, S5); it was reduced for the other two (S1 and S6), even in the interactive conditions.

These observations support the idea that the hyper-articulation characterizing Lombard speech concerns both the lips and the tongue and does not pertain to visible movements only. Despite this general tendency, there seem to be some slight inter-individual differences in the enhancement of these movements, particularly for LP and tongue displacement, which may reflect different strategies in making use of the visual modality to improve intelligibility.

Responses to the following questions can provide a second line of evidence:

Q4. Are visible movements more enhanced in noisy situations when the speaker can be seen by a speech partner (AV condition)?

Q5. Do speakers increase their vocal effort less in noisy situations when both audible and visible information are available (AV), compared to when they can only be heard by a speech partner (AO)?

Q6. Are barely visible movements comparably enhanced in noisy situations when the speaker can be seen vs only heard by a speech partner (AV and AO conditions)?

As expected, five speakers (S2, S3, S4, S5, S6) demonstrated more enhanced lip (visible) movements with NE in the AV condition. However, one speaker (S1) followed the same tendency as in his acoustic modifications and demonstrated more amplified lip movements in the AO condition.

Acoustic modifications related to an increase in vocal effort were indeed reduced in the AV condition for two speakers (S1 and S6). However, two other speakers (S2 and S5) showed the opposite behavior, namely greater acoustic modifications in the AV condition. The two remaining speakers (S3 and S4) did not show any significant difference between the two interaction conditions in their acoustic modifications in response to NE.

Finally, as regards less visible movements of the tongue between interaction conditions, speakers S1 and S6 again demonstrated more enhanced tongue movements with NE in the AO condition, whereas S2, S3 and S5 applied more enhanced movements in the AV condition. S4 enhanced the within-dispersion of tongue dorsum movements more in the AV condition.

To sum up, these observations do not appear to support the general idea that all speakers make use of the visual modality to improve their speech intelligibility in auditorily perturbed communication conditions (Hyp3). Instead, and as already suggested by the individual data in Hazan and Kim (2013), our results support the existence of speaker-specific strategies:

- Some speakers (like our participant S1) may not use the visual modality to improve their intelligibility (Hyp2). They may simply adapt to noise by "expanding sonority," that is, increasing their vocal loudness (Beckman *et al.*, 1992). An amplification of their articulatory movements (both visible and less visible) accompanies this main adaptation, but this may be related to acoustic modifications rather than an active strategy to improve the clarity of visible speech cues. This global shouting strategy follows the predictions of the hyper- and hypo-articulation (H&H) theory (Lindblom, 1990). It was more pronounced in the AO condition, when only audible information is available, than in the AV condition, in which redundant or complementary information is conveyed through the visual channel.
- Some speakers (like our participants S2, S3, S4, and S5, or like the participants in the Fitzpatrick *et al.*, 2015, study), appear to make active use of the visual modality to improve their intelligibility in noisy environments (Hyp3): they enhance their visible articulatory movements, particularly LP, in noisy conditions. They also enhance their visible articulatory movements more in the AV condition than in the AO condition. However, for the four subjects of this study, a greater increase in vocal intensity still accompanied these enhanced articulatory movements in the AV condition, although the H&H theory would instead predict that information enhancement in the visual domain could enable the speaker to limit the increase in his auditory effort even more. The more probable explanation may be that this greater increase in vocal intensity is not actually intended by the speakers but is simply an acoustic consequence of the increased lip radiation, directly related to the amplified lip opening.
- Finally, some speakers (like our participant S6) may "play" with both modalities and make the best use of their complementarity to improve their intelligibility in noise (Hyp2 and Hyp3): thus, S6 increased his vocal effort and amplified acoustic cues more in the AO condition, whereas he amplified LP cues and visible inter-vowel contrasts more in the AV condition than in the AO condition.

C. Limitations and future directions

The increased vocal effort observed in AO compared to AV conditions can be interpreted as (1) an active strategy to enhance audible cues when only that modality is available to convey information; (2) the regulation of speech efforts to meet listeners' needs, that is, as compensation for the "perturbation" in the communicative interaction induced by the lack of visual modality (H&H theory); or (3) a reflection of the fact that, in our protocol, the AO condition was defined by the experimenter turning her back to the speaker, whereas such a condition was simulated in other studies by adding a removable screen or curtain between the participants (Fitzpatrick *et al.*, 2015; Hazan and Kim, 2013), or having the interlocutors wear visors (Aubanel *et al.*, 2012).

Furthermore, less enhanced articulatory movements in the AO (last) condition may be interpreted as (1) an active strategy to enhance visible cues when the visual modality is available, or (2) a fatigue effect, since the protocol was quite long, with a fixed order for the experimental conditions, in which the AO condition is the last one.

We cannot rule out these alternative interpretations of our results, which would stem from our experimental protocol. However, the fact that not all our participants demonstrated greater vocal effort and/or less enhanced articulatory movements in the AO condition argues in favor of a negligible impact of these experimental limitations (fixed order of conditions + experimenter turning her back on the participants). Nevertheless, future studies on that topic would help clarify these results.

It should be noted that the task used in this experiment was constrained by the electromagnetic articulatory measurements and therefore involved a limited degree of communicative interaction. Since the experiment focused on segment intelligibility, the speech material was also limited to CVC target words differing only in the vowel or initial consonant. These experimental choices may lead to reduced speech adaptation with NE, compared to a more realistic communicative situation, or on the contrary, to the development of speech adaptation strategies focusing on the discrimination of the varying segment and on visual cue enhancement. It would therefore be necessary, in a future study, to explore whether the adaptive strategies reported in this article are still observed in more realistic conditions of communicative interaction, with more varied and common speech material.

Further studies should be conducted using perceptual assessment of the produced speech material in the various communicative contexts in order to test whether the speaker adjustments in production were successful, in other words, if they increased speech intelligibility in the auditory, visual, and AV domains. Furthermore, the inter-individual differences in speech adaptation observed in this study could be further investigated through a large-scale study involving participants differing by controlled anatomical or psychological factors (e.g., gender, lip shape, perceptual acuity, lip-reading abilities, etc.). This would shed light on the possible factors influencing the adoption of different adaptive strategies in adverse communication conditions.

Finally, we observed here that none of our six participants varied his lip opening movements independently of variations in voice intensity. This leads us to assume that, at least for these six participants, increasing voice intensity may be primarily controlled by increasing lip radiation, rather than by increasing vocal effort at the glottal level. We could even assume that, for some speakers, this may be the main goal of speech hyper-articulation in noisy conditions, rather than improving visual intelligibility. To further investigate this idea, acoustic simulations of lip radiation from our articulatory measurements could be conducted in order to estimate the impact of lip articulatory modifications on voice intensity and to compare these simulations with our actual measurement of voice intensity. It would also be useful, in a future study, to record the electroglottographic signal in addition to lip articulation and voice intensity, in order to explore the correlation between the variations in LA, amplitude of the glottal vibration, and voice intensity. This would enable a better understanding of the relationship between articulatory and glottal efforts, and whether hyper-articulation in adverse conditions may actually be an efficient strategy to increase audibility, as well as provide visible speech cues, without straining one's voice.

V. CONCLUSION

A general tendency was observed in the six speakers regarding the influence of communicative interaction: All the speakers modified their speech production significantly in noisy surroundings, not only in interactive but also in NI conditions. The modifications concerned not only parameters that are directly related to voice intensity, but also vowel- and consonant-specific descriptors and inter-vocalic contrasts. Articulatory modifications concerned not only the visible lip movements but also the less visible tongue movements. Overall, greater speech modifications were observed in interactive conditions.

On the other hand, the six speakers demonstrated different ways of adapting to the available sensory modalities of interaction: as expected, four of them enhanced their visible articulatory movements with NE more in the AV condition than in the AO condition. However, one participant showed the opposite behavior. The final participant applied an intermediate strategy, enhancing acoustic cues more in the AO condition and amplifying LP cues and visible inter-vowel contrasts more in the AV condition.

These results further support the idea that the Lombard effect is not simply an automatic regulation of one's own voice intensity but also a listener-oriented adaptation, which aims at improving speech intelligibility. On the other hand, they do not support the claim that speakers make deliberate use of the visual modality to improve their speech intelligibility in noisy conditions. In reality, only some speakers appear to do so.

ACKNOWLEDGMENTS

We thank our volunteer subjects for their participation, as well as Silvain Gerber for fruitful discussions of the statistical analysis, Gabrielle Richard for her help with the data acquisition, and Zofia Laubitz for copy-editing the paper.

[1]The term "active" refers here to a strategy that is controlled to directly impact intelligibility, as opposed to "passive," which would apply to a strategy that is a by-product of another strategy.

Alexanderson, S., and Beskow, J. (**2014**). "Animated Lombard speech: Motion capture, facial animation and visual intelligibility of speech produced in adverse conditions," Comput. Speech Lang. **28**, 607–618.

Amazi, D. K., and Garber, S. R. (**1982**). "The Lombard sign as a function of age and task," J. Speech Lang. Hear. Res. **25**, 581–585.

Arciuli, J., Simpson, B. S., Vogel, A. P., and Ballard, K. J. (**2014**). "Acoustic changes in the production of lexical stress during Lombard speech," Lang. Speech **57**, 149–162.

Aubanel, V., Cooke, M., Foster, E., Lecumberri, M. L. G., and Mayo, C. (**2012**). "Effects of the availability of visual information and presence of competing conversations on speech production," in *Proceedings of Interspeech 2012*, Portland, OR, pp. 2033–2036.

Auer, E. T., and Bernstein, L. E. (**2007**). "Enhanced visual speech perception in individuals with early-onset hearing impairment," J. Speech Lang. Hear. Res. **50**, 1157–1165.

Beckman, M. E., Edwards, J., and Fletcher, J. (**1992**). "Prosodic structure and tempo in a sonority model of articulatory dynamics," in *Papers in Laboratory Phonology II: Segment, Gesture, Prosody*, edited by G. J. Docherty and D. R. Ladd (Cambridge University Press, Cambridge, UK), pp. 68–86.

Beňuš, Š., Reichel, U. D., and Šimko, J. (**2015**). "F0 discontinuity as a marker of prosodic boundary strength in Lombard speech," in *Proceedings of Interspeech 2015*, Dresden, Germany, pp. 953–957.

Bernstein, L. E., Auer, E. T., and Takayanagi, S. (**2004**). "Auditory speech detection in noise enhanced by lipreading," Speech Commun. **44**, 5–18.

Bernstein, L. E., Tucker, P. E., and Demorest, M. E. (**2000**). "Speech perception without hearing," Percept. Psychophys. **62**, 233–252.

Bicevskis, K., de Vries, J., Green, L., Heim, J., Božič, J., D'Aquisto, J., Fry, M., Sadlier-Brown, E., Tkachman, O., Yamane, N., and Gick, B. (**2016**). "Effects of mouthing and interlocutor presence on movements of visible vs non-visible articulators," Can. Acoust. **44**, 17–24.

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.

Bond, Z. S., and Moore, T. J. (**1990**). "A note on loud and Lombard speech," in *First International Conference on Spoken Language Processing*, DTIC Document No. 969-972, http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA346202 (Last viewed 21/08/2018).

Bosker, H. R., and Cooke, M. (**2018**). "Talkers produce more pronounced amplitude modulations when speaking in noise," J. Acoust. Soc. Am. **143**, EL121–EL126.

Bourne, T., Garnier, M., and Samson, A. (**2016**). "Physiological and acoustic characteristics of the male music theatre voice," J. Acoust. Soc. Am. **140**, 610–621.

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

Chung, V., Mirante, N., Otten, J., and Vatikiotis-Bateson, E. (**2005**). "Audiovisual processing of Lombard speech," in *AVSP-2005*, pp. 55–56.

Conrad, R. (**1977**). "Lip-reading by deaf and hearing children," Br. J. Educ. Psychol. **47**, 60–65.
Cooke, M., King, S., Garnier, M., and Aubanel, V. (**2014**). "The listening talker: A review of human and algorithmic context-induced modifications of speech," Comput. Speech Lang. **28**, 543–571.
Cooke, M., and Lu, Y. (**2010**). "Spectral and temporal changes to speech produced in the presence of energetic and informational maskers," J. Acoust. Soc. Am. **128**, 2059–2069.
Davis, C., and Kim, J. (**2001**). "Repeating and remembering foreign language words: Implications for language teaching systems," Artif. Intell. Rev. **16**, 37–47.
Davis, C., Sironic, A., and Kim, J. (**2006**). "Perceptual processing of audio-visual Lombard speech," in *Proceedings of the 11th Australasian International Conference on Speech Science & Technology*, Auckland, New Zealand, pp. 248–252.
DePaulo, B. M., and Coleman, L. M. (**1986**). "Talking to children, foreigners, and retarded adults," J. Pers. Soc. Psychol. **51**, 945–959.
Dodd, B. (**1977**). "The role of vision in the perception of speech," Perception **6**, 31–40.
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.
Erber, N. P. (**1975**). "Auditory-visual perception of speech," J. Speech Hear. Disord. **40**, 481–492.
Fitzpatrick, M., Kim, J., and Davis, C. (**2015**). "The effect of seeing the interlocutor on auditory and visual speech production in noise," Speech Commun. **74**, 37–51.
Freed, B. F. (**1981**). "Foreigner talk, baby talk, native talk," Int. J. Sociol. Lang. **28**, 19–40.
Garnier, M. (**2008**). "May speech modifications in noise contribute to enhance audio-visible cues to segment perception?," in *Proceedings of AVSP'08*, Moreton Island, Australia, pp. 95–100.
Garnier, M., Bailly, L., Dohen, M., Welby, P., and Lœvenbruck, H. (**2006a**). "An acoustic and articulatory study of Lombard speech: Global effects on the utterance," in *Proceedings of Interspeech 2006*, Pittsburgh, PA, pp. 17–22.
Garnier, M., Dohen, M., Lœvenbruck, H., Welby, P., and Bailly, L. (**2006b**). "The Lombard effect: A physiological reflex or a controlled intelligibility enhancement?," in *Proceedings of the 7th ISSP*, Ubatuba, Brazil, pp. 255–262.
Garnier, M., and Henrich, N. (**2014**). "Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?," Comput. Speech Lang. **28**, 580–597.
Garnier, M., Henrich, N., and Dubois, D. (**2010**). "Influence of sound immersion and communicative interaction on the Lombard effect," J. Speech Lang. Hear. Res. **53**, 588–608.
Garnier, M., Ménard, L., and Richard, G. (**2012**). "Effect of being seen on the production of visible speech cues: A pilot study on Lombard speech," in *Proceedings of Interspeech 2012*, Portland, OR, pp. 611–614.
Geumann, A. (**2001**). "Vocal intensity: Acoustic and articulatory correlates," in *Proceedings of the 4th Conference on Motor Control*, Nijmegen, the Netherlands, pp. 1–4.
Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., and Yunusova, Y. (**2010**). "Lip movement exaggerations during infant-directed speech," J. Speech Lang. Hear. Res. **53**, 1529–1542.
Hazan, V., Grynpas, J., and Baker, R. (**2012**). "Is clear speech tailored to counter the effect of specific adverse listening conditions?," J. Acoust. Soc. Am. **132**, EL371–EL377.
Hazan, V., and Kim, J. (**2013**). "Acoustic and visual adaptations in speech produced to counter adverse listening conditions," in *Proceedings of AVSP'13*, Annecy, France, pp. 93–98.
Henriques, R. N., and van Lieshout, P. (**2013**). "A comparison of methods for decoupling tongue and lower lip from jaw movements in 3D articulography," J. Speech Lang. Hear. Res. **56**, 1503–1516.
Huber, J. E., and Chandrasekaran, B. (**2006**). "Effects of increasing sound pressure level on lip and jaw movement parameters and consistency in young adults," J. Speech Lang. Hear. Res. **49**, 1368–1379.
Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., and Johnson, K. (**1999**). "Formants of children, women, and men: The effects of vocal intensity variation," J. Acoust. Soc. Am. **106**, 1532–1542.
Jiang, J., Alwan, A., Keating, P. A., Auer, E. T., and Bernstein, L. E. (**2002**). "On the relationship between face movements, tongue movements, and speech acoustics," EURASIP J. Adv. Signal Process. **11**, 1174–1188.
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. (**1999**). "The Lombard effect: A reflex to better communicate with others in noise," in *Proceedings of ICASSP*, Phoenix, AR, pp. 2083–2086.
Kim, J., and Davis, C. (**2014**). "Comparing the consistency and distinctiveness of speech produced in quiet and in noise," Comput. Speech Lang. **28**, 598–606.
Kim, J., Davis, C., Vignali, G., and Hill, H. (**2005**). "A visual concomitant of the Lombard reflex," in *Proceedings of AVSP'05*, Vancouver, Canada, pp. 17–21.
Kim, J., Sironic, A., and Davis, C. (**2011**). "Hearing speech in noise: Seeing a loud talker is better," Perception **40**, 853–862.
Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.
Lienard, J. S., and Di Benedetto, M. G. (**1999**). "Effect of vocal effort on spectral properties of vowels," J. Acoust. Soc. Am. **106**, 411–422.
Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Springer, Dordrecht), pp. 403–439.
Lindblom, B. E., and Sundberg, J. E. (**1971**). "Acoustical consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Am. **50**, 1166–1179.
Lombard, E. (**1911**). "Le signe de l'élévation de la voix" ["The sign of raising the voice"], Ann. Mal. Oreille Larynx Nez Pharynx **37**, 101–119.
Lu, Y., and Cooke, M. (**2008**). "Speech production modifications produced by competing talkers, babble, and stationary noise," J. Acoust. Soc. Am. **124**, 3261–3275.
Lu, Y., and Cooke, M. (**2009**). "Speech production modifications produced in the presence of low-pass and high-pass filtered noise," J. Acoust. Soc. Am. **126**, 1495–1499.
MacLeod, A., and Summerfield, Q. (**1987**). "Quantifying the contribution of vision to speech perception in noise," Br. J. Audiol. **21**, 131–141.
Manabe, K., Sadr, E. I., and Dooling, R. J. (**1998**). "Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): Differential reinforcement of vocal intensity and the Lombard effect," J. Acoust. Soc. Am. **103**, 1190–1198.
Ménard, L., Leclerc, A., and Tiede, M. (**2014**). "Articulatory and acoustic correlates of contrastive focus in congenitally blind adults and sighted adults," J. Speech Lang. Hear. Res. **57**, 793–804.
Ménard, L., Trudeau-Fisette, P., Côté, D., and Turgeon, C. (**2016**). "Speaking clearly for the blind: Acoustic and articulatory correlates of speaking conditions in sighted and congenitally blind speakers," PloS One **11**, e0160088.
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.
Patel, R., and Schell, K. W. (**2008**). "The influence of linguistic content on the Lombard effect," J. Speech Lang. Hear. Res. **51**, 209–220.
Perkell, J. S., Denny, M., Lane, H., Guenther, F., Matthies, M. L., Tiede, M., Vick, J., Zandipour, M., and Burton, E. (**2007**). "Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users," J. Acoust. Soc. Am. **121**, 505–518.
Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R., and Kearney, J. K. (**1989**). "Inhibiting the Lombard effect," J. Acoust. Soc. Am. **85**, 894–900.
Pittman, A. L., and Wiley, T. L. (**2001**). "Recognition of speech produced in noise," J. Speech Lang. Hear. Res. **44**, 487–496.
Reisberg, D., Mclean, J., and Goldfield, A. (**1987**). "Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli," in *Hearing by Eye: The Psychology of Lip-Reading*, edited by B. Dodd and R. Campbell (Lawrence Erlbaum, London), pp. 97–113.
Robert-Ribes, J., Schwartz, J. L., Lallouache, T., and Escudier, P. (**1998**). "Complementarity and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise," J. Acoust. Soc. Am. **103**, 3677–3689.
Rostolland, D. (**1982a**). "Phonetic structure of shouted voice," Acta Acust. **51**, 80–89.
Rostolland, D. (**1982b**). "Acoustic features of shouted voice," Acta Acust. **50**, 118–125.
Savariaux, C., Perrier, P., and Orliaguet, J. P. (**1995**). "Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production," J. Acoust. Soc. Am. **98**, 2428–2442.
Schulman, R. (**1989**). "Articulatory dynamics of loud and normal speech," J. Acoust. Soc. Am. **85**, 295–312.

Siegel, G. M., Pick, H. L., Olsen, M. G., and Sawin, L. (**1976**). "Auditory feedback on the regulation of vocal intensity of preschool children," Dev. Psychol. **12**, 255–261.
Šimko, J., Beňuš, Š., and Vainio, M. (**2016**). "Hyperarticulation in Lombard speech: Global coordination of the jaw, lips and the tongue," J. Acoust. Soc. Am. **139**, 151–162.
Sinnott, J. M., Stebbins, W. C., and Moody, D. B. (**1975**). "Regulation of voice amplitude by the monkey," J. Acoust. Soc. Am. **58**, 412–414.
Sodersten, M., Ternstrom, S., and Bohman, M. (**2005**). "Loud speech in realistic environmental noise: Phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers," J. Voice **19**, 29–46.
Stanton, B. J., Jamieson, L. H., and Allen, G. D. (**1988**). "Acoustic-phonetic analysis of loud and Lombard speech in simulated cockpit conditions," in *Proceedings of ICASSP*, New York, NY, pp. 331–334.
Sumby, H., and Pollack, I. W. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Summerfield, Q. (**1992**). "Lipreading and audio-visual speech perception," Philos. Trans. R. Soc. London B Biol. Sci. **335**, 71–78.
Tasko, S. M., and McClean, M. D. (**2004**). "Variations in articulatory movement with changes in speech task," J. Speech Lang. Hear. Res. **47**, 85–100.
Ternström, S., Södersten, M., and Bohman, M. (**2002**). "Cancellation of simulated environmental noise as a tool for measuring vocal performance during noise exposure," J. Voice **16**, 195–206.
Tonkinson, S. (**1994**). "The Lombard effect in choral singing," J. Voice **8**, 24–29.
Turner, G., Roach, L., and de Jonge, R. (**2016**). "Lip contact pressure while talking in background noise," Perspect. ASHA Spec. Interest Groups **1**, 5–14.
Vainio, M., Suni, A., Arnhold, A., Raitio, T., Seijo, H., Järvikivi, J., Aalto, D., and Alku, P. (**2012**). "Effect of level and type of noise on focus related prosody," in *The Listening Talker*, Edinburgh, Scotland, p. 85.
Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.
Vatikiotis-Bateson, E., Barbosa, A. V., Chow, C. Y., Oberg, M., Tan, J., and Yehia, H. C. (**2007**). "Audiovisual Lombard speech: Reconciling production and perception," in *Proceedings of AVSP'07*, Hilvarenbeek, the Netherlands, pp. 45–50.
Wassink, A. B., Wright, R. A., and Franklin, A. D. (**2007**). "Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers," J. Phon. **35**, 363–379.
Welby, P. (**2006**). "Intonational differences in Lombard speech: Looking beyond F0 range," in *Proceedings of Speech Prosody 2006*, Dresden, Germany, pp. 763–766.
Yehia, H., Rubin, P., and Vatikiotis-Bateson, E. (**1998**). "Quantitative association of vocal-tract and facial behavior," Speech Commun. **26**, 23–43.
Zeiliger, J., Serignat, J. F., Autresserre, D., and Meunier, C. (**1994**). "BD_Bruit, une base de données de parole de locuteurs soumis à du bruit" ["BD_Bruit, a speech database from speakers exposed to noise"], in *Proceedings of the Xèmes Journées d'Etude sur la Parole*, Grenoble, France, pp. 287–290.