

Understanding the Lombard Effect for Mandarin: Relation Between Speech Recognition Thresholds and Acoustic Parameters

Trends in Hearing

Volume 29: 1–13

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/23312165251324266

journals.sagepub.com/home/tia



Fei Chen¹ , Changjie Pan¹, Hongmei Hu², Sabine Hochmuth^{2,3}, Birger Kollmeier² and Anna Warzybok²

Abstract

The present work quantifies the Lombard effect across native speakers of Mandarin Chinese using the Matrix sentence test, which is optimized for precisely assessing speech recognition thresholds (SRTs) in noise. Specifically, we studied the effects of speaker gender, fundamental frequency (F0), formant frequencies (F1 and F2), the duration and rate of voiced segments, and frequency-specific energy redistribution characterized by alpha ratio and speech-weighted signal-to-noise ratio (swSNR) on the recognition of Mandarin in plain and Lombard speech. The Mandarin Chinese matrix test was recorded with plain and Lombard speech from 11 native-Mandarin speakers. SRTs in stationary noise were measured with native-Mandarin, normal-hearing listeners. Results showed that on average, Mandarin Lombard speech was more intelligible than Mandarin plain speech for both female and male speakers, and the Mandarin Lombard gain of female speakers was larger than that of males. In addition, various acoustic analyses involving all speakers showed that (a) only swSNR was significantly correlated with the SRT of the Mandarin plain speech; (b) most acoustic measures were significantly correlated with the SRT of the Mandarin Lombard speech; and (c) alpha ratio and swSNR were significantly correlated with the SRT Lombard gain. In addition, a gender effect was found in the correlational analysis between acoustic parameters and SRT as well as Lombard gain in SRT. The findings highlight the impact of increased high-frequency energy on the observed Lombard gain in Mandarin speech, whereas the changes in individual acoustic parameters (e.g., F0 and F1) appear to play only a minor role.

Keywords

Lombard effect, speech recognition threshold, alpha ratio, speech-weighted signal-to-noise ratio (swSNR), SRT Lombard gain

Received: 1 July 2024; revised: 24 January 2025; accepted: 9 February 2025

Introduction

In noisy environments, most people change the way they speak. This kind of speech modification in the presence of noise was discovered by Lombard and has since become known as the Lombard effect (Lombard, 1911). Typically, the intelligibility of Lombard speech is higher than that of plain speech (produced in quiet) at equal signal-to-noise ratio (SNR) (e.g., Cooke et al., 2014; Dreher and O'Neill, 1957; Lu and Cooke, 2008, 2009; Van Summers et al., 1988). Generally, Lombard speech can be characterized by a collection of acoustic and phonetic modifications (see Cooke et al., 2014 for a review). Compared with plain speech uttered in

¹Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China

²Medizinische Physik and Cluster of Excellence "Hearing4all", Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

³Department of Otolaryngology, Head and Neck Surgery, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Corresponding author:

Anna Warzybok, Medizinische Physik and Cluster of Excellence

"Hearing4all", Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany.

Email: anna.warzybok-oetjen@uni-oldenburg.de

Data Availability Statement included at the end of the article



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

quiet, Lombard speech exhibits a higher F0 and a higher first formant frequency (F1), an increased (Junqua, 1993) or decreased (Pisoni et al., 1985) second formant frequency (F2), a flattened spectral tilt (Cooke and Lu, 2010; Lu and Cooke, 2009), an increased intensity (Cooke and Lu, 2010; Kleczkowski et al., 2017), and a greater vowel duration (Alghamdi et al., 2018; Junqua and Anglade, 1990). These acoustic and phonetic modifications may lead to increased intelligibility of Lombard speech. The gender difference in the Lombard effect has also been explored in several studies. Junqua (1993) reported that in the presence of additive babble noise, the Lombard speech of female speakers was more intelligible than that of male speakers, while the plain speech of male speakers was more intelligible than that of female speakers.

To date, the phenomena of Lombard effects have been observed in different languages (e.g., Kleczkowski et al., 2017; Uma Maheswari et al., 2021). However, most of the early work focused on Lombard speech in English (e.g., Dreher and O'Neill, 1957; Pisoni et al., 1985; Van Summers et al., 1988), and the studies in Chinese languages were limited (e.g., Ng and Tsang, 2019; Tang et al., 2017a, 2017b; Yang et al., 2022; Zhao and Jurafsky, 2009). The present work further studied the Lombard effect on Mandarin speech, and the unique contributions of the present work are reflected in the following aspects:

1. Mandarin, as a tonal language, differs in several aspects from non-tonal languages (e.g., Chen et al., 2013). It is unclear how important acoustic cues (e.g., F0) for Mandarin speech perception contribute to the perceptual advantage (if any) of Mandarin Lombard speech. The findings of this work contribute to understanding the language effect or specificity of Lombard speech.
2. Earlier work studied the Lombard gain under specifically selected SNR conditions (e.g., Cooke et al., 2014; Lee et al., 2017). That is, the Lombard gain was only reported for fixed SNR level(s). Psychometric curves are commonly used to reflect the influence of external noise interference on speech perception, whereas the noise (or SNR) level varies over a wide range. The psychometric curve and its characteristics, including slope and speech recognition threshold (SRT, the SNR level at which a speech perception rate of 50% is found), may provide a more comprehensive view of the relation between speech perception and noise interference (e.g., Plomp and Mimpen, 1979). Furthermore, SRT measurements make it possible to avoid the ceiling- or floor effect that is typically observed in fixed SNR measurements. An established, efficient method for SRT determination is the Matrix sentence test (see Kollmeier et al., 2015 for a review) which is available in more than 20 languages including Mandarin (Hu et al., 2018). Therefore, the present

work used SRT determined by the Matrix sentence test to characterize the influence of noise on speech comprehension and Lombard gain in Mandarin. Due to the high comparability of carefully constructed Matrix sentence tests across languages, this study has the potential to serve as a reference study for comparing the Lombard effect across different languages.

3. In addition to the impacts of F0 and SNR conditions, many other cues (e.g., low-frequency region (Fogerty and Chen, 2014), formants (Han and Chen, 2017), etc.) also demonstrate specific contributions to Mandarin speech perception, which may differ from those contributions observed in non-tonal languages. The present work carried out an acoustic analysis to investigate the changes of speech cues (e.g., F0 and energy distribution across frequency bands) between Mandarin Lombard and plain speech.
4. Most of the previous studies reported the perceptual gain of Lombard speech, but they did not specifically investigate the effect of speaker on the observed gain. The present work recorded the Mandarin Lombard/plain speech from a total of 11 native-Mandarin speakers, and studied the Lombard gain for each individual speaker. In addition, correlation analysis was carried out between existing acoustic analyses (e.g., F0, F1, F2, and energy distribution across frequency bands) and Lombard gain. The findings of this work provide insights into important acoustic correlates of the perceptual gain of Mandarin Lombard speech.

Methods

Speech Material and Recording

This work used the Mandarin Chinese matrix sentence test (CMNmatrix, Hu et al., 2018) as testing materials, which was compatible with matrix tests for other major languages (Kollmeier et al., 2015 for a review) and was developed based on the ICRA guidelines (Akeroyd et al., 2015). The CMNmatrix is based on two- or three-syllabic words that are frequently used in spoken Mandarin speech. It contains 10 each of the following: names, verbs, numerals, nouns, and adjectives. These 50 words are combined into grammatically correct but semantically neutral and unpredictable sentences (e.g., “张伟看见一个普通的水壶” in Mandarin, or “Zhangwei saw one ordinary kettle” in English).

For the purpose of this study, the speech material of the CMNmatrix was recorded from 11 native-Mandarin speakers (5 females and 6 males, denoted in the following as SF1-5 and SM1-6, respectively) using both plain and Lombard speech. All speakers spoke fluent standard Mandarin. One hundred sentences (ten base lists of ten sentences) of the CMNmatrix were recorded from each speaker in both plain and Lombard speaking styles (each base list containing all 50 words). The 100 sentences were divided into 10 blocks

of 10 sentences each, and the plain and Lombard blocks were presented in alternating order. The order of the sentences within a block was randomized. The recording took place in a double-walled, sound-attenuated booth fulfilling ISO 8253-3 (ISO 8253-3, 2012), using a Neumann 184 microphone with a cardioid characteristic (Georg Neumann GmbH, Berlin, Germany) and a Fireface UC soundcard (with a sampling rate of 44,100 Hz and resolution of 16 bits). The recording procedure generally followed the procedures of Alghamdi et al. (2018). A Mandarin-native speaker and a phonetician participated in the recording session and listened to the sentences to control the pronunciations, intonation, and speaking rate. During the recording, the speaker was instructed to read the sentence presented on a frontal screen. The speaker was asked to utter the sentences in an intermediate speaking rate, which was facilitated by a progress bar on the screen. In case of any mispronunciation or change in the intonation, the speaker was asked via the screen to repeat the sentence again, and on average, each sentence was recorded twice. In Lombard conditions the speaker was regularly asked (every three to five sentences) via a prompt to repeat a sentence, to keep the speaker in the Lombard communication situation. For the plain-speech recording blocks, the speakers were asked to pronounce the sentences with natural intonation and accentuation. Furthermore, the speakers were asked to keep the speaking effort constant and to avoid any exaggerated pronunciations that could lead to unnatural speech cues. For the Lombard speech recording blocks, speakers were instructed to imagine a conversation to another person in a pub-like situation. Throughout the whole recording session, the speakers wore headphones (Sennheiser HDA200) that reproduced the speaker's audio signal. The level of the playback of the speakers' speech was adjusted individually (as described by Alghamdi et al., 2018). Briefly, the perception of the own voice with and without headphones was compared. Therefore, the speaker wore the headphone over one ear, leaving the other ear uncovered. The speaker was asked to speak while the playback of his/her voice was presented at gradually increasing levels through the headphones. The speaker's task was to indicate the level at which the signals were perceived equally loud in both ears. This level was recorded individually and used as the presentation level of the playback during recordings. In the Lombard condition, the stationary speech-shaped noise ICRA1 (Dreschler et al., 2001) presented at a level of 80 dB SPL was mixed with the speaker's playback. Calibration of the noise signal was done using a Brüel & Kjær (B&K) 4153 artificial ear, a B&K 4134 0.5-inch microphone, a B&K 2669 preamplifier, and a B&K 2610).

Previous studies showed that the masker level of 80 dB SPL induced a robust Lombard speech without the danger of inducing hearing damage (Alghamdi et al., 2018). The recording session lasted about one hour thirty minutes in total. This included some breaks between the recording blocks. For half of the recording session the talker was exposed to the

masking noise. By that, the exposure to the noise (considering the presentation level of the noise and its duration) was substantially below the recommended exposure limits.

The sentences were cut manually from the recording, high-pass filtered (60 Hz cut-off frequency) and set to the average root-mean-square level of the original speech material of the CMNmatrix (Hu et al., 2018). Then, the best version of each sentence was chosen by native-Mandarin speakers so that the selected sentences were comparable in terms of pronunciation, tempo, and intonation.

The Lombard and plain speech with different speakers is available at Zenodo (Hu et al. 2022).

Listening Experiment: Listeners

Thirteen native-Mandarin-speaking listeners (9 male and 4 female) aged between 21 and 34 years (mean 24.4 years) participated in the listening experiment. All listeners were normal-hearing with pure-tone thresholds of 20 dB hearing level or better at audiometric octave frequencies between 125 and 8,000 Hz. All listeners provided written informed consent, as approved by the Ethics Committee of the Southern University of Science and Technology. Listeners received an hourly compensation for their participation.

Listening Experiment: Measurement Procedure

The measurements took place in a double-walled and sound-attenuating booth. Speech recognition measurements were administered using the Oldenburg Measurement Applications software (HörTech gGmbH, Oldenburg, Germany). The speech and noise were presented monaurally (to a single ear) through free-field-equalized Sennheiser (HDA300) headphones (Sennheiser, Germany) via a laptop and external sound card MAYA 22. The measurement setup was calibrated to dB SPL using a Brüel & Kjær artificial ear 4152, type 4144 microphone, and a 2250 sound-level meter.

The sound pressure level of the noise was kept constant at 65 dB SPL; thus the SNR was determined by the speech-signal level. The noise signal started 500 ms before and ended 500 ms after the presentation of each sentence, and was gated with 50 ms rising and falling ramps (using a Hann window). An adaptive procedure suitable for concurrent estimation of SRT and slope was applied (Brand and Kollmeier, 2002). This procedure converged to a so-called pair of compromise target values corresponding to 20% and 80% correct responses.

Each listener attended 22 listening conditions [=11 talkers \times 2 speaking styles (plain and Lombard)]. In order to obtain sufficient accuracy of SRT and slope estimates, test lists of 30 sentences were used for each listening condition. Therefore, three randomly selected test lists of 10 sentences were concatenated. The order of both the sentences within a list and the listening conditions was randomized. Note that

the 22 listening conditions were tested on two consecutive days (11 conditions per day) using the same ear. The listeners were asked to repeat the words they heard and the test instructor marked the correct responses in the software. The word-correct rates were recorded during the adaptive procedure for the estimation of SRT and slope. Prior to the first measurement, two lists of 20 sentences and one list of 20 sentences were conducted on day one and day two, respectively, to account for the training effect (Hu et al., 2018). For training purposes, the original version of the CMNmatrix was used (Hu et al., 2018), and the feedback of correct responses were provided to the listeners. Each listening condition (with 30 sentences) took about 6 min, and a 5-min break was given every 30 min to avoid listening fatigue. Each listener took about 1.5 h to finish the experimental procedure each day.

Acoustic Analyses

Motivated by earlier studies (e.g., Alghamdi et al., 2018), seven acoustic parameters were used in this work to study the acoustic difference between the Mandarin plain and Lombard speech, including alpha ratio, F0, F1, F2, the mean length of voiced segments, the number of continuous voiced segments per second, and speech-weighted SNR (swSNR).

The four acoustic parameters of alpha ratio, F0, the mean length of voiced segments, and the number of continuous voiced segments per second were obtained using the publicly available openSMILE toolkit (<http://audeering.com/technology/opensmile/>, last viewed on 01 August 2021), which was used in earlier studies to analyze the acoustic parameters of Lombard speech (e.g., Alghamdi et al., 2018). Specifically, the alpha ratio was defined as the ratio between the speech energies of 1 to 15 kHz and 50 to 1,000 Hz frequency regions computed in the openSMILE toolkit (Sundberg and Nordenberg, 2006). The trajectories of F1 and F2 of sentence materials were extracted using PRAAT, and the PRAAT script to implement the above processing is available at (<https://phonetics.linguistics.ucla.edu/facilities/acoustic/SWS.txt>, last viewed on 17 October 2024). Note that the PRAAT script extracted three formants (i.e., F1, F2, and F3) but, consistent with early studies (e.g., Alghamdi et al., 2018), only the first two formant frequencies were analyzed in this study. The maximum frequencies in formant tracker were set to 5,500 and 5,000 Hz for female and male speech, respectively. The mean values of F0, F1, and F2 correspond to the mean across the matrix sentences. The openSMILE toolkit was also used to extract the mean length of voiced segments and the number of continuous voiced segments per second.

swSNR is another measure characterizing the spectral difference across the plain and Lombard speech. In contrast to the alpha ratio, however, it considers spectral characteristics of both speech and noise signals. swSNR is a measure of an effective SNR taking into account the relative contributions of different regions of the frequency

spectrum to speech intelligibility (Greenberg et al., 1993). To calculate the speech-weighted SNR, speech and noise signals were first divided into octave bands. Then, the SNRs in each frequency band were weighted according to their contribution to speech intelligibility. The frequency-weighting function represented the average speech and was taken from Table III of the speech intelligibility index (SII) standard (ANSI, 1997). For calculations of the swSNR the actual speech signals were replaced by a stationary speech-simulating noise generated from all the sentences (separately for each talker and speaking style, i.e., plain and Lombard). In this way, the noise had the same long-term frequency spectrum as the speech signal. The intrinsic variance of speech would require very long signals and/or multiple predictions with different speech signals in order to make the variability of the prediction independent of the specific speech sample. This is a commonly used method in calculation and comparisons of speech and noise signals (Beutelmann et al., 2010). Therefore, for each talker and speaking style only one value of swSNR is provided.

For statistical analysis, the mean values (utterances-per-speaker, or all utterances across all speakers) of those acoustic parameters were computed to characterize the Mandarin plain and Lombard speech. Paired *t*-test was used to compare the difference (significance level was set at $P < .05$) between paired SRT scores (or acoustic parameters) of plain and Lombard speech.

Results

Lombard Effect of Male and Female Mandarin Speech

Figure 1(a) shows the mean SRT scores (with corresponding standard error of the mean) of all female and male speakers under the plain and Lombard conditions. Statistical significance was determined by using SRT as the dependent variable and using the gender (female and male) and speaking style (plain and Lombard) as two within-subject factors. Two-way repeated-measures analysis of variance (ANOVA, via SPSS toolkit) indicated a significant effect of gender ($F_{1, 12} = 183.15$, $P < .005$), a significant effect of speaking style ($F_{1, 12} = 20.92$, $P < .005$), and a non-significant interaction between these two within-subject variables ($F_{1, 12} = 0.96$, $P = .345$). As seen in Figure 1(a), on average, plain speech resulted in SRTs of -15.8 ± 0.6 dB and -13.9 ± 0.8 dB for female and male speakers, respectively. The mean SRTs for Lombard speech were -17.7 ± 0.7 dB and -14.5 ± 0.7 dB for female and male speakers, respectively. Post-hoc paired *t*-tests revealed statistically significant ($P < .05$, $\alpha = .05$) difference between plain and Lombard speech for female speaker's group ($P < .0001$) and male speaker's group ($P = .01$). Figure 1(b) shows the empirical Lombard gain (i.e., the difference between the SRTs of plain and Lombard speech) for female and male speakers. Paired *t*-test



Figure 1. (a) Speech recognition thresholds (SRTs) averaged across speakers and listeners under all listening conditions. Asterisks denote a significant ($P < 0.05$) difference between the SRTs of the paired conditions. (b) Empirical Lombard gains for female and male speakers across all listeners. The error bars denote ± 1 standard error of the mean.

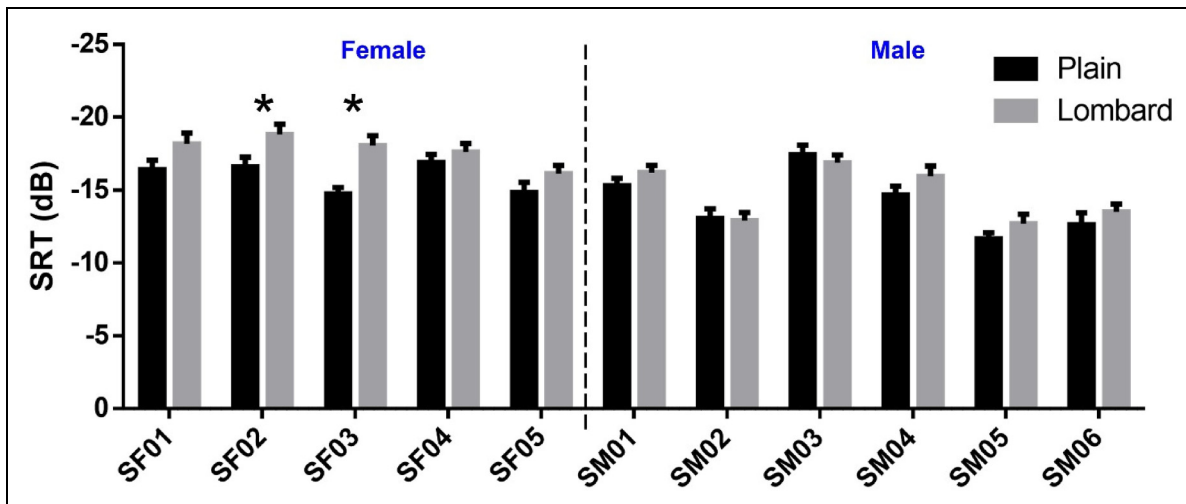


Figure 2. SRTs with corresponding standard errors of the Mandarin plain and Lombard speech of all speakers. Asterisks denote that the SRTs difference between plain and Lombard speech are significant ($P < 0.0045$) or non-significant ($P > 0.0045$), respectively. ‘SF’ and ‘SM’ denote female and male speakers, respectively.

shows that the female speakers group demonstrates a significantly ($P < .05$) larger Lombard gain than the male speaker’s group, that is, 1.9 dB versus 0.5 dB.

Figure 2 shows the mean SRTs with corresponding standard error of the mean (across listeners) of the Mandarin plain and Lombard conditions for each speaker. Mauchly’s test indicated that the assumption of sphericity was violated for the main effect of speaker ($\chi^2(54) = 95.09$, $P < .005$). Therefore, degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity. All effects were significant, that is, $P < .05$. There were significant main effects of speaking style (plain and Lombard) ($F_{1, 11} = 68.06$, $P < .05$), and of the speaker ($F_{4.83, 53.08} = 66.97$, $P < .05$) on SRT. Also, the interaction between speaking style and the speaker was statistically significant ($F_{10, 110} = 3.80$, $P = .006$). Multiple paired

comparisons with Bonferroni correction were run between the SRT values of plain and Lombard speech of each speaker. The Bonferroni-corrected statistical significance level was set at $P < .0045$ ($\alpha = .05$). Paired t -tests revealed statistically significant Lombard gain for 2 out of 11 speakers (indicated by the symbol “*” in Fig. 2). That is, for two female speakers (i.e., SF02 and SF03), there were statistically significant ($P < .0045$) differences in SRTs measured with plain and Lombard speech, $P = .0008$ and $P < .0001$ for SF02 (-16.5 ± 2.3 dB vs. -18.7 ± 2.4 dB) and SF03 (-14.6 ± 1.5 dB vs. -17.8 ± 2.4 dB), respectively.

Note that both the SRT and slope of the psychometric curve reflect the effect of noise interference on speech comprehension. While this work primarily focused on the SRT difference between the Mandarin plain and Lombard speech,

experimental results also provided slopes of the psychometric curves corresponding to different speakers. Specifically, the mean slope for plain speech (11 speakers) was 11.4%/dB, ranging from 9.3%/dB (F01) to 14.9%/dB (M05). Corresponding mean slopes for female and male speakers were $10.5 \pm 1.4\%$ /dB and $12.1 \pm 1.8\%$ /dB, respectively. Slightly steeper slopes were observed for the Lombard speech with a mean slope of $12.8 \pm 1.8\%$ /dB and $13.8 \pm 0.8\%$ /dB for female and male speakers, respectively.

Acoustic Analysis for Mandarin Lombard and Plain Speech

Figure 3 shows the acoustic parameters, introduced in section “Acoustic analyses,” analyzed for each speaker group (i.e., all, female, male). The mean alpha ratio was significantly (all $P < .05$) lower for the plain speech than for the Lombard speech for conditions with female (−10.9 vs. −7.4), male (−11.5 vs. −9.7) and all speakers (−11.3 vs. −8.7), and the magnitude of this difference was statistically larger ($P < .05$) for female speakers (3.5) than for male speakers (1.8). The same trends were observed for the swSNR-based results. On average, the mean of swSNR was significantly (all $P < .05$) higher for Lombard speech than for plain speech for conditions with female- (4.3 dB vs. 3.2 dB), male- (2.8 dB vs. 2.0 dB) and all-speakers (3.5 dB vs. 2.5 dB), and this difference was slightly higher for female speakers (1.1 dB) than for male speakers (0.8 dB). The average F0 and F1 were significantly (all $P < .05$) increased from plain speech to the Lombard speech for conditions with female (F0: 37.0 Hz vs. 39.3 Hz, F1: 683.3 Hz vs. 714.7 Hz), male (F0: 28.2 Hz vs. 30.7 Hz, F1: 611.5 Hz vs. 625.9 Hz) and all speakers (F0: 32.2 Hz vs. 34.6 Hz, F1: 644.1 Hz vs. 666.3 Hz), but there was no significant ($P > .05$) difference between F2 s of plain and Lombard speech for conditions with female (1,774.8 Hz vs. 1,772.3 Hz), male (1,579.7 Hz vs. 1,571.2 Hz) and all speakers (1,668.4 Hz vs. 1,662.6 Hz). For Lombard speech, the duration of voiced segments was significantly (all $P < .05$) increased, and the rate of voiced segments was significantly (all $P < .05$) reduced for conditions with female (duration: 0.23 s vs. 0.28 s, rate: 3.3 vs. 2.8), male (duration: 0.18 s vs. 0.22 s, rate: 4.0 vs. 3.6) and all speakers (duration: 0.21 s vs. 0.25 s, rate: 3.7 vs. 3.3).

Relationship Between Acoustic Parameters and SRT

To test the relation between acoustic parameters and the observed SRT values, a correlation analysis was performed between the mean values of the selected acoustic parameter and the mean SRT values under the “all” condition (i.e., involving all 11 speakers), or the male speaker or female speaker condition, respectively. Table 1 gives the Pearson correlation coefficients of the 7 acoustic parameters under both the plain and Lombard conditions under all, male

and female conditions. For plain speech involving all speakers, only swSNR yielded a significant ($P < .05$) correlation coefficient $r = -0.85$ [see scatter plot in Figure 4(a)] as opposed to other measures (all $|r| < 0.6$, all $P > .05$). The correlation coefficients of alpha ratio, F0, F1, and F2 ranged from $r = -0.51$ to $r = -0.56$, and those of voiced segment duration and rate were -0.35 and -0.30 , respectively. For plain speech involving 6 male speakers, again only swSNR yielded a significant ($P < .05$) correlation coefficient $r = -0.85$ (see scatter plot in Figure 4(b)). For plain speech involving 5 female speakers, none of those acoustic parameters significantly correlated with the mean SRT values, that is, all $P > .05$. Note that a gender effect is seen in Table 1, that is, the correlation analysis applied to the condition of male or female speakers produces clearly different results. For the condition of male speakers, only the alpha ratio and swSNR yielded significant ($P < .05$) correlation coefficients, that is, -0.92 and -0.93 , respectively. On the other hand, for female speakers, only F1 significantly correlated with the SRT of the Lombard speech.

Notably, the correlation coefficients increase when involving all speakers, that is, performing correlation analysis between the mean values of the selected acoustic parameter and the mean SRT values of Lombard speech. Specifically, swSNR also yielded the highest correlation coefficient $r = -0.96$ ($P < .05$), followed by alpha ratio ($r = -0.81$, $P < .05$). Other acoustic parameters (except voiced segment duration) were also significantly ($P < .05$) correlated with the SRTs of Lombard speech, ranging from $r = 0.61$ (voiced segment rate) to $r = -0.75$ (F0). Hence, swSNR was significantly—and best—correlated with the SRT of the Mandarin plain or Lombard speech when involving all speakers. Figure 4 shows the scatter plots between the swSNR measure and the SRT value of the Mandarin plain or Lombard speech under all, male and female conditions.

Table 1 also shows the results when using a selected acoustic parameter to correlate with the empirical SRT Lombard gain. The empirical SRT Lombard gain was determined as the difference between the SRT values of plain and Lombard speech, that is, $SRT_{\text{plain}} - SRT_{\text{Lombard}}$. For the acoustic parameters, the Lombard gain was similarly defined as the difference between the plain and Lombard speech. It is seen in Table 1 that both swSNR ($r = 0.71$) and alpha ratio ($r = 0.78$) yielded the highest correlation coefficients (both $P < .05$). Both are higher than the correlation coefficients of other acoustic parameters, ranging from 0.36 (F0) to 0.66 (F2), with only F2 being significant. Furthermore, a gender effect can be observed, as both alpha ratio and F2 yield non-significant ($P > .05$) correlation coefficients under the male condition, and swSNR yields non-significant ($P > .05$) correlation coefficient under both male and female conditions. Figure 4 also shows the scatter plots between the swSNR measure and the empirical SRT Lombard gain under three gender conditions (i.e., all, male, and female).

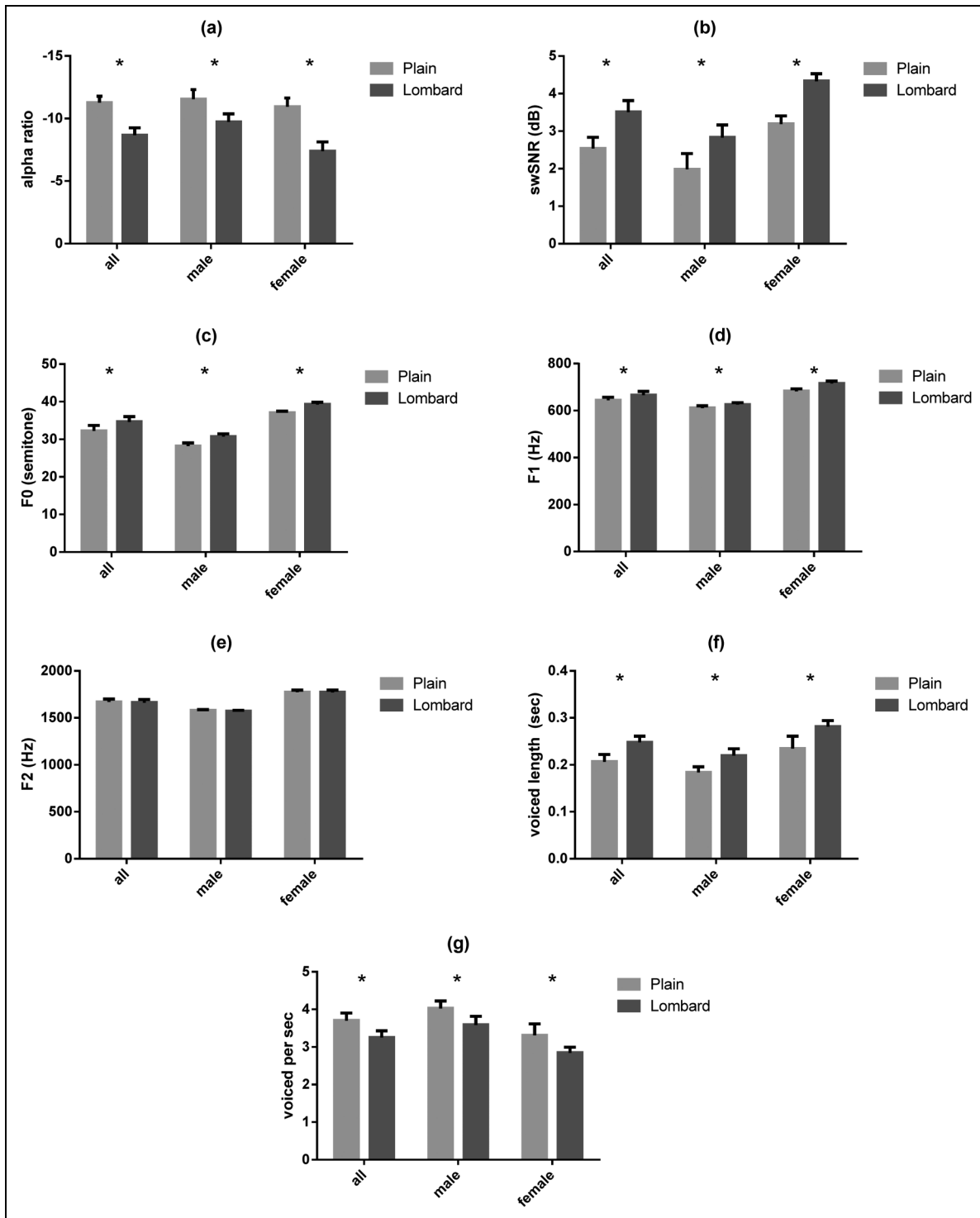


Figure 3. Mean values of acoustic parameters for Mandarin plain and Lombard speech of different speaker groups. Panels (a), (b), (c), (d), (e), (f) and (g) are for alpha ratio, swSNR, F0, F1, F2, duration of voiced segments, and rate of voiced segments, respectively. Asterisks denote a significant ($P < 0.05$) difference between the SRTs of the paired conditions. The error bars denote ± 1 standard error of the mean. 'all', 'male' and 'female' denote the comparisons involving all speakers, male speakers, and female speakers, respectively.

Table 1. Pearson Correlation Coefficients (r) Between the Selected Acoustic Parameter and the SRT or Lombard Gain in SRT Under Three Gender Conditions (i.e., All, Male, and Female). Correlation Coefficients in Black Denote That There was a Significant ($P < .05$) Correlation Between the Selected Acoustic Parameter and the SRT or SRT Lombard Gain Across Speakers. Note That the Difference of the Respective Acoustic Parameter Between Plain and Lombard Speech was Correlated With the SRT Lombard Gain.

	SRT (Plain)			SRT (Lombard)			SRT Lombard gain		
	All	Male	Female	All	Male	Female	All	Male	Female
alpha ratio	−0.56 ($P=.071$)	−0.74 ($P=.096$)	−0.06 ($P=.924$)	−0.81 ($P=.003$)	−0.92 ($P=.009$)	−0.19 ($P=.761$)	0.78 ($P=.005$)	0.40 ($P=.433$)	−0.97 ($P=.005$)
swSNR	−0.85 ($P<.001$)	−0.85 ($P=.033$)	−0.52 ($P=.365$)	−0.96 ($P<.001$)	−0.93 ($P=.007$)	−0.76 ($P=.133$)	0.71 ($P=.024$)	0.24 ($P=.643$)	0.82 ($P=.091$)
F0	−0.55 ($P=.082$)	−0.18 ($P=.731$)	−0.78 ($P=.118$)	−0.75 ($P=.008$)	−0.09 ($P=.865$)	−0.59 ($P=.292$)	0.36 ($P=.275$)	0.51 ($P=.306$)	0.24 ($P=.696$)
F1	−0.55 ($P=.083$)	−0.53 ($P=.282$)	0.31 ($P=.616$)	−0.71 ($P=.015$)	−0.67 ($P=.147$)	0.88 ($P=.049$)	0.46 ($P=.155$)	0.30 ($P=.559$)	0.04 ($P=.951$)
F2	−0.51 ($P=.107$)	−0.02 ($P=.977$)	−0.40 ($P=.499$)	−0.73 ($P=.011$)	−0.14 ($P=.797$)	−0.23 ($P=.706$)	0.66 ($P=.028$)	0.91 ($P=.011$)	0.43 ($P=.467$)
Voiced segment duration	−0.35 ($P=.294$)	−0.30 ($P=.563$)	0.04 ($P=.946$)	−0.57 ($P=.065$)	−0.36 ($P=.478$)	0.70 ($P=.186$)	0.36 ($P=.280$)	0.38 ($P=.455$)	0.66 ($P=.227$)
Voiced segment rate	0.30 ($P=.367$)	0.13 ($P=.802$)	−0.14 ($P=.825$)	0.61 ($P=.048$)	0.47 ($P=.346$)	−0.68 ($P=.212$)	0.40 ($P=.223$)	0.41 ($P=.421$)	0.64 ($P=.245$)

Discussion

The Lombard effect has been widely noted for English, Dutch, Polish, and other non-tonal languages (e.g., Bosker and Cooke, 2020; Cooke and Lecumberri, 2012; Kłeczkowski et al., 2017), but a systematic acoustic and perceptual investigation of the Lombard effect for a tonal language like Mandarin is still lacking. Due to the language difference, the mechanisms of understanding Mandarin and English speech have many differences, which have been extensively reported in earlier studies (e.g., Fogerty and Chen, 2014). The current work studied the Mandarin Lombard effect by human speech recognition measurements with the Matrix sentence test which has been optimized to efficiently and precisely determine the SRT in a way compatible with more than 20 languages (Kollmeier et al., 2015). In addition to empirical measurements, this work investigated several factors influencing speech recognition of plain and Lombard speech, including the gender effect, speaker variance, as well as acoustic and phonetic parameters.

In contrast to earlier studies of the Lombard effect that investigated speech intelligibility at fixed levels or SNR, the current study assessed speech recognition thresholds, measured with the CMNmatrix test, to characterize the respective average speech recognition performance across a group of individuals with normal hearing. The advantage of the SRT measurement is that it provides a representative score (i.e., SNR level at 50% word correct rate) that is not influenced by the selected SNRs of the test conditions. Although our experiments study the Lombard effect for Mandarin, some findings of the present work were largely consistent with earlier results. First, the Mandarin Lombard speech provided a better speech recognition, that is, a lower SRT, which clearly demonstrated the perceptual advantage of Mandarin Lombard speech and was thus similar to the findings for non-tonal languages (Lu and Cooke, 2008; Pittman and Wiley, 2001; van

Summers et al. 1988). Second, this work showed substantial differences in SRTs for plain and Lombard speech, as well as in Lombard benefit across the speakers. Generally, recordings from female speakers resulted in better intelligibility and Lombard gain was larger for female than male speakers, in agreement with findings of earlier studies for non-tonal languages (e.g., Ferguson, 2004; Junqua, 1993).

The present study analyzed seven acoustic parameters to illustrate differences between the Mandarin plain and Lombard speech. Note that of the many important acoustic parameters that could have been compared, this work specifically analyzed the frequency spectrum and voicing information (F0, F1, and F2). This choice of observables was mainly due to the importance of voiced segments on Mandarin speech perception. For instance, lexical tone identification (carried out in voiced segments) is important for Mandarin speech comprehension in noise (e.g., Chen et al., 2014), and vowels carry more perceptual importance than consonants (e.g., Chen et al., 2013). Considering the changes in the energy distribution across frequencies between plain and Lombard speech, the outcomes of this study confirm the effects demonstrated in the past for non-tonal languages (Cooke et al., 2014; Lu and Cooke, 2008, 2009). This indicates that the shift in energy towards higher frequencies for Lombard speech is a characteristic feature change in the speech signal for both tonal and non-tonal languages. The current work showed a significant increase of F0 and F1 in Lombard speech relative to plain speech, which is again in agreement with previous studies analyzing phonetic changes in Lombard speech for non-tonal languages (Alghamdi et al., 2018; Lu and Cooke, 2008). The significant differences in acoustic parameters observed between plain and Lombard speech apply to both female and male speakers. Notably, for parameters characterizing the frequency-specific energy distribution (i.e., alpha ratio and swSNR), the magnitude of these differences was larger for

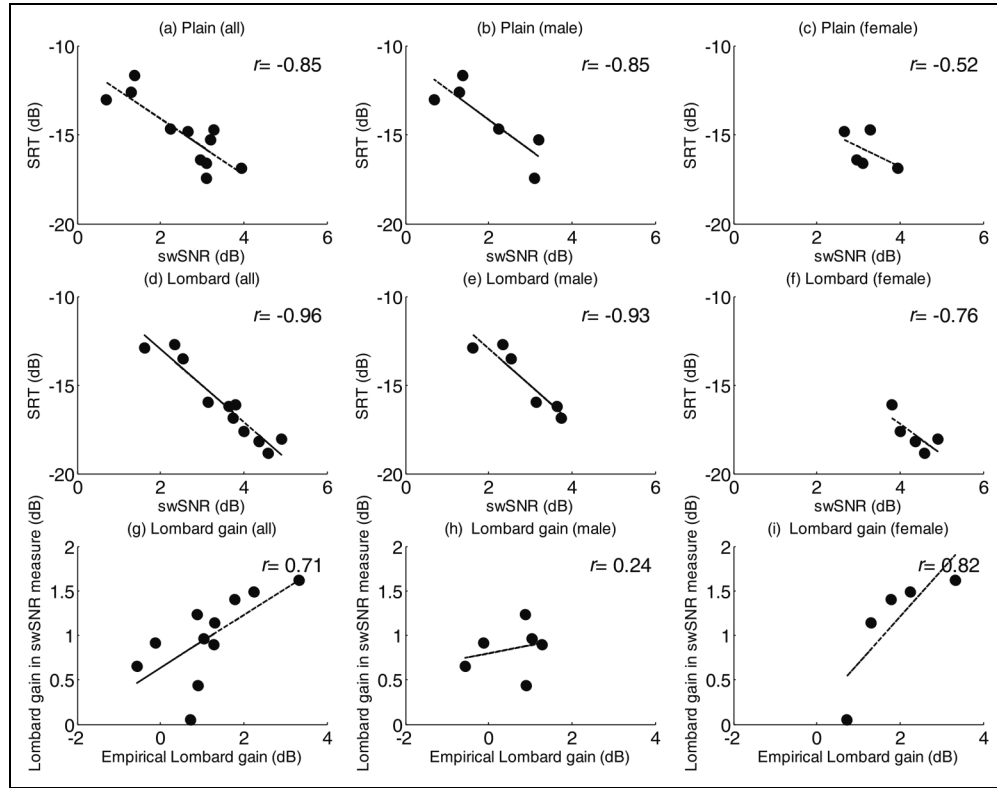


Figure 4. Scatter plots between SRT and swSNR for the plain speech of (a) all speakers, (b) male speakers, and (c) female speakers, and for the Lombard speech of (d) all speakers, (e) male speakers, and (f) female speakers. The lowest row depicts scatter plots between empirical Lombard gain in SRT and Lombard gain in swSNR for (g) all speakers, (h) male speakers, and (i) female speakers.

female speakers than for male speakers. However, such a gender effect was not observed for F0, duration, or the rate of voiced segments.

The next step examined how the changes in the acoustic and phonetic parameters were related to the speech intelligibility of Mandarin plain and Lombard speech, and whether they can sufficiently predict the Lombard benefit. The shift in energy to higher frequency regions for Lombard speech, as characterized via the alpha ratio and swSNR, correlated well with the Lombard speech SRTs and Lombard benefit. This demonstrates that, as for non-tonal languages (Lu and Cooke, 2008), the spectral properties of the speech and noise are crucial for speech intelligibility and can be characterized by rather simple acoustic measures. The low correlation values between Lombard gain in SRT for male speakers and the alpha ratio or swSNR can be explained by the fact that Lombard gain was nearly absent for male speakers, averaging only 0.5 dB. However, what differs across non-tonal and tonal languages is the contribution of changes in F0 to intelligibility. In most of the previous studies for non-tonal languages, the contribution of F0 changes to improved intelligibility of Lombard speech was negligible (Lu and Cooke, 2009). In the current study, a significant correlation between F0 and the SRT of Lombard speech was observed (i.e., $r = -0.75$ in Table 1). However, this significant correlation disappeared

when the analysis was conducted separately by gender. Therefore, the observed correlation between F0 and SRT is likely to reflect the impact of gender on SRT. However, since a larger change in F0 is observed between both genders than within each gender group, this gender effect on SRT cannot be separated from an independent effect of F0 on SRT. The impacts of increasing F0 and middle- and high-frequency speech energy on improving speech recognition could be partially attributed to the better perception of pitch information. Future investigations could extend the analysis of F0-related aspects by considering the F0 dynamics, which may differ between tonal and non-tonal languages. Remaining acoustic and phonetic parameters considered here, that differed significantly between plain and Lombard speech, that is, the increased duration of voiced segments and a low rate of voiced syllables for Lombard speech, were not significantly related to Lombard gain in terms of SRT. Also for non-tonal languages, little or no role of the durational increases seen in Lombard speech was related to the Lombard intelligibility benefit (Cooke et al., 2014).

A high variability in SRT across the speakers in plain and Lombard speech was observed within this study. Considering plain speech only, the SRT variability across the speakers could be partially explained by swSNR (with $r = -0.85$). All remaining acoustic and phonetic measures were

statistically not related to the SRTs of plain speech and are therefore weak measures to characterize the speaker-specific SRTs for speech produced in quiet. Interestingly, a significant ($P < .05$, except voiced segment duration) and moderately high correlation (from $r = -0.71$ to $r = -0.96$) with the SRT of Lombard speech was found. Hence, parameters considered here may provide useful insights into the design of speech-intelligibility indices for predicting Lombard speech intelligibility. This could be extended in future studies by applying auditory models to our data in order to better understand the underlying principles for understanding plain and Lombard speech, as well as the variability across speakers. Candidates are the SII model (ANSI3.5, 1997), glimpsing model (Cooke, 2006), or automatic speech-recognizer-based models such as the framework of auditory discrimination experiments (Schädler et al., 2015, 2016). The SII relies on a decomposition of speech and noise into spectral components, and has proven itself to be a simple and yet powerful predictor for speech recognition in non-tonal languages such as in English or German (Hargus & Gordon-Salant, 1995; Lopez-Poveda et al., 2017; Wardenga et al., 2015). SII type models have been optimized and adapted in the past to predict speech recognition for tonal languages such as Mandarin. This adaptation is achieved by the notion of the so-called language- and gender-specific band importance function (Chen et al., 2016; Du et al., 2018). However, it is not clear if the differences between the frequency band importance functions arise from differences across languages or rather reflect and characterize a particular talker. Furthermore, speech produced with normal, raised or loud vocal effort is also characterized by different frequency band importance functions based on the ANSI standards (1997). Again, it is not clear if these functions are valid for a specific language and talker only or if the differences across the speech with different vocal effort can be transferred to other languages. These questions could be addressed by SII simulations of the SRT data described in this paper. Furthermore, simulations with the automatic speech-recognizer-based model (Schädler et al., 2015, 2016) could contribute to better understanding of the importance of different features for speech recognition of plain and Lombard speech as well as Lombard gain. Speech recognition can be simulated using different representation of the speech and noise signals, for example, using the Log-Mel-Spectrograms, or separable Gabor filterbank with only spectral, only temporal or both spectro-temporal features. Comparisons of simulated SRTs obtained with different representation of speech and noise signals may provide valuable information about which features are crucial to account for the effects observed in the present study.

Conclusions

The present work examined the perceptual benefit of the Lombard effect for Mandarin, and related the subjective data to the acoustic analyses of the speech signals.

Particularly, acoustic factors such as F0, F1, F2, duration and rate of voiced segments were considered, as well as the distribution of the energy across the frequencies (i.e., alpha ratio and swSNR). The main conclusions are:


1. Using SRT to study the Mandarin Lombard gain across genders and speakers, the findings demonstrated that the Lombard effect for Mandarin was more evident for female speakers, since the empirical Lombard gain of female speakers was significantly larger than that of male speakers.
2. For Mandarin Lombard speech, acoustic analyses demonstrate that F0 and F1, but not F2, were increased in comparison to plain speech. Similarly, the alpha ratio and swSNR increased. The Mandarin Lombard speech showed increased voiced-segment duration and decreased voice-segment rate. These findings were largely consistent with those previously reported for English speech materials (e.g., Alghamdi et al., 2018).
3. For the Mandarin Lombard speech, among the acoustic parameters examined in this work, alpha ratio, swSNR, F0, F1, F2 and voiced-segment rate were well correlated with the SRT scores of Lombard speech. For plain speech, the SRT was only significantly correlated with swSNR. However, a gender effect was observed for the correlation analysis: Many significant correlation relationships disappeared for male speakers in comparison to female speakers indicating a strong influence of gender on SRT which, however, cannot be separated from an independent effect of F0 on SRT.
4. Among all factors examined, it was found that both measures reflecting the distribution of the energy across the frequencies, that is, alpha ratio and swSNR, were well correlated with the Mandarin Lombard gain. Again, a gender effect was found when correlating alpha ratio or swSNR with Lombard gain. These findings highlight the importance of the spectral changes to produce SRT changes via the Lombard effect, whereas the other features appear to play a smaller role.

Acknowledgments

English-language services were provided by stels-ol.de.

ORCID iDs

Fei Chen  <https://orcid.org/0000-0002-6988-492X>

Anna Warzybok  <https://orcid.org/0000-0002-4298-4605>

Statements and Declarations

Author Contributions

F. Ch. contributed to the design of the study and acoustic-phonetic analyses. A. W. contributed to the design of the study, analyzed speech intelligibility measurement including empirical data and

objective measures. S.H. prepared the speech material, made recordings, prepared the speech material for speech intelligibility measurements and implemented it in the measurement software. Ch. P. conducted the listening experiments and contributed to the acoustic and phonetic analyses, H. H. contributed to the preparation of the speech material, assessed the quality of the recordings and selected the sentences for speech intelligibility measurements, B. K. contributed to the design of the study. All authors discussed the results and contributed to the manuscript.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (grant numbers 2020 M-0021, 62371217), Basic and Applied Basic Research Foundation of Guangdong Province (grant number 2022B1515120056), and Deutsche Forschungsgemeinschaft (grant numbers 390895286, 415895050).

Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability

The Lombard and plain speech with different speakers (version 1.0) is freely available at Zenodo (<https://doi.org/10.5281/zenodo.7063030>, Hu et al., 2022).

References

- Akeroyd, M. A., Arlinger, S., Bentler, R. A., Boothroyd, A., Dillier, N., Dreschler, W. A., Gagné, J., Lutman, M., Wouters, J., Wong, L., & Kollmeier, B. (2015). International collegium of rehabilitative audiology (ICRA) recommendations for the construction of multilingual speech tests. *International Journal of Audiology*, 54(Suppl. 2), 17–22. <https://doi.org/10.3109/14992027.2015.1030513>
- Alghamdi, N., Maddock, S., Marxer, R., Barker, J., & Brown, G. J. (2018). A corpus of audio-visual Lombard speech with frontal and profile views. *Journal of the Acoustical Society of America*, 143(6), EL523–EL529. <https://doi.org/10.1121/1.5042758>
- ANSI. (1997). *American National Standard Methods for the calculation of the speech intelligibility index*. American National Standards Institute.
- Beutelmann, R., Brand, T., & Kollmeier, B. (2010). Revision, extension, and evaluation of a binaural speech intelligibility model. *Journal of the Acoustical Society of America*, 127(4), 2479–2497. <https://doi.org/10.1121/1.3295575>
- Bosker, H. R., & Cooke, M. (2020). Enhanced amplitude modulations contribute to the Lombard intelligibility benefit: Evidence from the Nijmegen corpus of Lombard speech. *Journal of the Acoustical Society of America*, 147(2), 721–730. <https://doi.org/10.1121/10.0000646>
- Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *Journal of the Acoustical Society of America*, 111(6), 2801–2810. <https://doi.org/10.1121/1.1479152>
- Chen, J., Huang, Q., & Wu, X. (2016). Frequency importance function of the speech intelligibility index for mandarin Chinese. *Speech Communication*, 83, 94–103. <https://doi.org/10.1016/j.specom.2016.07.009>
- Chen, F., Wong, L. L. N., & Hu, Y. (2014). Effects of lexical tone contour on mandarin sentence intelligibility. *Journal of Speech, Language, and Hearing Research*, 57(1), 338–345. [https://doi.org/10.1044/1092-4388\(2013/12-0324](https://doi.org/10.1044/1092-4388(2013/12-0324)
- Chen, F., Wong, L. L. N., & Wong, Y. W. E. (2013). Assessing the perceptual contributions of vowels and consonants to mandarin sentence intelligibility. *Journal of the Acoustical Society of America*, 134(2), EL178–EL184. <https://doi.org/10.1121/1.4812820>
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, 119(3), 1562–1573. <https://doi.org/10.1121/1.2166600>
- Cooke, M., & Lecumberri, M. L. (2012). The intelligibility of Lombard speech for non-native listeners. *Journal of the Acoustical Society of America*, 132(2), 1120–1129. <https://doi.org/10.1121/1.4732062>
- Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *Journal of the Acoustical Society of America*, 128(4), 2059–2069. <https://doi.org/10.1121/1.3478775>
- Cooke, M., Mayo, C., & Villegas, J. (2014). The contribution of durational and spectral changes to the Lombard speech intelligibility benefit. *Journal of the Acoustical Society of America*, 135(2), 874–883. <https://doi.org/10.1121/1.4861342>
- Dreher, J. J., & O'Neill, J. (1957). Effects of ambient noise on speaker intelligibility for words and phrases. *Journal of the Acoustical Society of America*, 29(12), 1320–1323. <https://doi.org/10.1121/1.1908780>
- Dreschler, W. A., Verschuure, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. International Collegium for Rehabilitative Audiology. *Audiology*, 40(3), 148–157.
- Du, Y., Shen, Y., Yang, H., Wu, X., & Chen, J. (2018). Measuring the band importance function for Mandarin Chinese with a Bayesian adaptive procedure. Proceedings of the Interspeech 2018, pp. 961–965. <https://doi.org/10.21437/interspeech.2018-1825>
- Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, 116(4), 2365–2373. <https://doi.org/10.1121/1.1788730>
- Fogerty, D., & Chen, F., 2014. Vowel spectral contributions to English and Mandarin sentence intelligibility. In: Proceeding of 15th Annual Conference of the International Speech Communication Association (InterSpeech), pp. 499–503.
- Greenberg, J. E., Peterson, P. M., & Zurek, P. M. (1993). Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. *Journal of the Acoustical*

- Society of America*, 94(5), 3009–3010. <https://doi.org/10.1121/1.407334>
- Han, Y. Y., & Chen, F. (2017). Relative contributions of formants to sentence intelligibility in mandarin Chinese. *Journal of the Acoustical Society of America*, 141(6), EL495–EL499. <https://doi.org/10.1121/1.4983747>
- Hargus, S. E., & Gordon-Salant, S. (1995). Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment. *Journal of Speech and Hearing Research*, 38(1), 234–243. <https://doi.org/10.1044/jshr.3801.234>
- Hu, H. M., Warzybok, A., Scharf, M., Kollmeier, B., Chen, F., & Hochmuth, S. (2022). Mandarin matrix recordings: Lombard and plain speech with different speaks (version 1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7063030>
- Hu, H. M., Xi, X., Wong, L. L. N., Hochmuth, S., Warzybok, A., & Kollmeier, B. (2018). Construction and evaluation of the mandarin Chinese matrix (CMNmatrix) sentence test for the assessment of speech recognition in noise. *International Journal of Audiology*, 57(11), 838–850. <https://doi.org/10.1080/14992027.2018.1483083>
- International Organization for Standardization. (2012). *Acoustics-audiometric test methods-part 3: Speech audiometry* (pp. 1–36). ISO 8253-3. ISO.
- Junqua, J. C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America*, 93(1), 510–524. <https://doi.org/10.1121/1.405631>
- Junqua, J. C., & Anglade, Y. (1990). Acoustic and perceptual studies of Lombard speech: Application to isolated-words automatic speech recognition. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (pp. 841–844).
- Kleczkowski, P., Zak, A., & Krol-Nowak, A. (2017). Lombard effect in Polish speech and its comparison in English speech. *Archives of Acoustics*, 42(4), 561–569. <https://doi.org/10.1515/aoa-2017-0060>
- Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology*, 54(sup2), 3–16. <https://doi.org/10.3109/14992027.2015.1020971>
- Lee, J. W., Ali, H., Ziaei, A., Tobey, E. A., & Hansen, J. H. L. (2017). The Lombard effect observed in speech produced by cochlear implant users in noisy environments: A naturalistic study. *Journal of the Acoustical Society of America*, 141(4), 2788–2799. <https://doi.org/10.1121/1.4979927>
- Lombard, E. (1911). The sign of the elevation of the voice. *Annals of the Diseases for Ear, Larynx, Nose, and Pharynx*, 37, 101–119.
- Lopez-Poveda, E. A., Johannesen, P. T., Pérez-González, P., Blanco, J. L., Kalluri, S., & Edwards, B. (2017). Predictors of hearing-aid outcomes. *Trends in Hearing*, 21, 233121651773052. <https://doi.org/10.1177/2331216517730526>
- Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *Journal of the Acoustical Society of America*, 124(5), 3261–3275. <https://doi.org/10.1121/1.2990705>
- Lu, Y., & Cooke, M. (2009). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication*, 51(12), 1253–1262. <https://doi.org/10.1016/j.specom.2009.07.002>
- Ng, M. L., & Tsang, G. (2019). The Lombard effect associated with Chinese male alaryngeal speech. *International Journal of Speech-Language Pathology*, 21(6), 584–592. <https://doi.org/10.1080/17549507.2018.1551932>
- Pisoni, D. B., Bernacki, R. H., Nusbaum, H. C., & Yuchtman, M. (1985). Some acoustic-phonetic correlates of speech produced in noise. In: Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (pp. 1581–1584).
- Pittman, A. L., & Wiley, T. L. (2001). Recognition of speech produced in noise. *Journal of Speech, Language, and Hearing Research*, 44, 487–496. [https://doi.org/10.1044/1092-4388\(2001/038\)](https://doi.org/10.1044/1092-4388(2001/038))
- Plomp, R., & Mimpfen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18(1), 43–52. <https://doi.org/10.3109/00206097909072618>
- Schädler, M. R., Warzybok, A., Ewert, S. D., & Kollmeier, B. (2016). A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. *Journal of the Acoustical Society of America*, 139(5), 2708–2722. <https://doi.org/10.1121/1.4948772>
- Schädler, M. R., Warzybok, A., Hochmuth, S., & Kollmeier, B. (2015). Matrix sentence intelligibility prediction using an automatic speech recognition system. *International Journal of Audiology*, 54(sup2), 100–107. <https://doi.org/10.3109/14992027.2015.1061708>
- 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. *Journal of the Acoustical Society of America*, 120(1), 453–457. <https://doi.org/10.1121/1.2208451>
- Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017a). Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech. *Journal of the Acoustical Society of America*, 142(5), 2823–2835. <https://doi.org/10.1121/1.5008372>
- Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017b). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *Journal of the Acoustical Society of America*, 142(2), 493–503. <https://doi.org/10.1121/1.4995998>
- Uma Maheswari, S., Shahina, A., & Nayeemulla Khan, A. (2021). Understanding Lombard speech: A review of compensation techniques towards improving speech based recognition systems. *Artificial Intelligence Review*, 54, 2495–2523. <https://doi.org/10.1007/s10462-020-09907-5>
- Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production:

- Acoustic and perceptual analyses. *Journal of the Acoustical Society of America*, 84(3), 917–928. <https://doi.org/10.1121/1.396660>
- Wardenga, N., Batsoulis, C., Wagener, K. C., Brand, T., Lenarz, T., & Maier, H. (2015). Do you hear the noise? The German matrix sentence test with a fixed noise level in subjects with normal hearing and hearing impairment. *International Journal of Audiology*, 54(Suppl. 2), 71–79. <https://doi.org/10.3109/14992027.2015.1079929>
- Yang, Y., Chen, X., Liu, Q., Tu, W., Chen, H., & Cai, L. (2022). Mandarin Lombard grid: A Lombard-grid-like corpus of standard Chinese. In: *Proceedings of the Interspeech* (pp. 3078–3082).
- Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, 37(2), 231–247. <https://doi.org/10.1016/j.wocn.2009.03.002>