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.*

Enhanced amplitude modulations contribute to the Lombard intelligibility benefit: evidence from the Nijmegen Corpus of Lombard Speech

Hans Rutger Bosker[a1] and Martin Cooke[b2]

[a]*Psychology of Language department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, PO Box 310, 6500 AH, Nijmegen, The Netherlands*

[b]*Language and Speech Laboratory, Universidad del País Vasco, calle Justo Vélez de Elorriaga 1, Vitoria, 01006, Spain*

Accepted for publication in *The Journal of the Acoustical Society of America*

January 10, 2020

[authors' accepted manuscript]

[1] Corresponding author. Tel.: +31 (0)24 3521 373. *E-mail address:* HansRutger.Bosker@mpi.nl
Also at: *Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands*
[2] Also at: *Ikerbasque (Basque Science Foundation), Bilbao, Spain*

ABSTRACT

Speakers adjust their voice when talking in noise, which is known as Lombard speech. These acoustic adjustments facilitate speech comprehension in noise relative to plain speech (i.e., speech produced in quiet). However, exactly which characteristics of Lombard speech drive this intelligibility benefit in noise remains unclear. This study assessed the contribution of enhanced amplitude modulations to the Lombard speech intelligibility benefit by demonstrating that (1) native speakers of Dutch in the Nijmegen Corpus of Lombard Speech (NiCLS) produce more pronounced amplitude modulations in noise vs. in quiet; (2) more enhanced amplitude modulations correlate positively with intelligibility in a speech-in-noise perception experiment; (3) transplanting the amplitude modulations from Lombard speech onto plain speech leads to an intelligibility improvement, suggesting that enhanced amplitude modulations in Lombard speech contribute towards intelligibility in noise. Results are discussed in light of recent neurobiological models of speech perception with reference to neural oscillators phase-locking to the amplitude modulations in speech, guiding the processing of speech.

Keywords: Lombard speech; speech in noise; amplitude modulations; prosody transplantation; neural entrainment.

I. INTRODUCTION

When communicating in noisy acoustic environments, human and non-human species typically adjust their vocalizations. One of the most salient modifications is an increase in vocalization amplitude in proportion to the noise level, ultimately attempting to maintain a favorable signal-to-noise ratio (SNR; Hotchkin & Parks, 2013; Luo et al., 2015). Humans also exhibit other adjustments to their speech when speaking in noise, such as slower speech rate, raised fundamental frequency (F0), and flatter spectral tilt (for an overview, see Cooke, King, et al., 2014). Together, these noise-induced modifications result in what is collectively known as Lombard speech (i.e., speech produced in noise; Lombard, 1911), in contrast to 'unmodified' plain speech (speech produced in quiet). Functionally, Lombard speech is more intelligible than plain speech when presented in noise, even after discounting intensity increases (Dreher & O'Neill, 1957; Pittman & Wiley, 2001; Summers et al., 1988). However, exactly which acoustic characteristics of Lombard speech contribute to this intelligibility benefit in noise is not well understood. The present study, introducing the Nijmegen Corpus of Lombard Speech (NiCLS; publicly available for download), assessed the contribution of enhanced amplitude modulations, suggesting that more pronounced amplitude modulations in Lombard speech aid intelligibility.

Some previous studies have targeted the acoustic correlates of the intelligibility benefit of Lombard speech in noise. Lu and Cooke (2009) assessed the contribution of changes in F0 and spectral tilt. They collected plain speech recordings and flattened the spectral tilt, increased the F0, or both by means of artificial signal processing techniques, thus matching the characteristics of Lombard speech. While flattening of spectral tilt contributed greatly to the intelligibility benefit of Lombard speech in (speech-shaped) noise, increasing F0 did not have a significant influence.

However, changes in spectral tilt alone could not fully account for the intelligibility of Lombard speech and, therefore, the authors speculated that other, perhaps durational, vocal modifications may contribute to intelligibility as well. This speculation was tested by Cooke, Mayo, and Villegas (2014). Since Lombard speech typically has a slower speech rate than plain speech, Cooke et al. applied durational modifications to plain speech (linear and nonlinear time warping via time alignment), as well as spectral changes at the global utterance level and to individual time frames. While the spectral modifications produced an increase in intelligibility (albeit still falling short of that of Lombard speech itself), the durational modifications did not increase intelligibility at all. This suggests that spectral modifications drive much of the Lombard speech intelligibility benefit, which was further corroborated by Godoy, Koutsogiannaki, and Stylianou (2014). They showed that Lombard speech consistently exhibits spectral energy boosting in an inclusive formant region, effectively increasing audibility. A Lombard-inspired artificial signal processing technique involving spectral shaping and audio-enhancement techniques (i.e., a combination of Lombard-like Spectral Shaping (SS) and dynamic range compression (DRC); SSDRC; Godoy et al., 2014) was demonstrated to increase intelligibility, as indicated by both an energy-based metric, the speech intelligibility index, and keywords correct scores.

However, one aspect of Lombard speech that has received little attention concerns how talkers adjust the temporal modulations of their speech when conversing in noise. Speech in its very nature is an acoustic signal that contains strong amplitude modulations, particularly in the 1-15 Hz range (Ding et al., 2017; Flinker et al., 2019; Steeneken & Houtgast, 1980; Varnet et al., 2017). Speech intelligibility greatly relies on these amplitude modulations, evident in the temporal envelope of speech (Drullman et al., 1994a; Shannon et al., 1995). In fact, enhancing the amplitude modulations in speech makes it more intelligible in noise (Koutsogiannaki & Stylianou, 2016), while filtering

amplitude modulations in the 1-9 Hz range out of the speech signal impairs intelligibility to a large degree (Drullman et al., 1994a, 1994b; Ghitza, 2012).

Electrophysiological studies have demonstrated that speech-envelope information evokes marked "envelope-following" neural responses in the auditory cortex (Peelle & Davis, 2012). This 'speech tracking' has been taken by current neurobiological models of speech perception (Ghitza, 2011; Giraud & Poeppel, 2012) to explain the robust contribution of amplitude modulations to speech intelligibility. Endogenous neural oscillators in the lower frequency range (delta: 1-4 Hz; theta: 4-8 Hz) are thought to phase-lock to the amplitude fluctuations in the input signal (Bosker, 2017; Doelling et al., 2014; Kösem et al., 2018). This neural tracking of the temporal envelope of speech is proposed to underlie successful speech-in-noise and speech-in-speech intelligibility. That is, the phase of brain oscillations is primarily aligned to the dynamics of the attended (vs. the ignored) speech (Ding & Simon, 2012; Kerlin et al., 2010). Moreover, some studies have claimed a causal link, suggesting that the greater the alignment of cortical oscillators to the temporal envelope of the attended signal, the greater its intelligibility (Golumbic, Ding, et al., 2013; Golumbic et al., 2012; Rimmele et al., 2015).

Based on these neurobiological models, Bosker and Cooke (2018) assessed whether speakers, potentially in an attempt to aid speech intelligibility, would also naturally produce more enhanced amplitude modulations when talking in a noisy acoustic environment. Using modulation spectra, they observed more pronounced amplitude modulations in the temporal envelope of Lombard speech compared to plain speech, as evidenced by greater power in the lower frequency range of the modulation spectra, across a collection of four different speech corpora. However, only production data were reported in the study by Bosker and Cooke (2018). As such, the contribution of this greater power in the modulation domain in Lombard speech to speech intelligibility in noise

remains unknown. Furthermore, only English corpora were analyzed in Bosker and Cooke (2018); thus, further cross-linguistic validation is called for.

The present study investigated the contribution of enhanced amplitude modulations in Lombard speech to intelligibility in noise by means of both production and perception experiments. First, Experiment 1 introduces the NiCLS corpus. Native speakers of Dutch were recorded producing Lombard sentences (produced while speech-shaped noise was presented over headphones) and matching plain speech sentences (the same sentences produced in quiet). Adopting the methods of Bosker and Cooke (2018), we expected to find greater power in the modulation spectrum of Dutch Lombard speech (vs. plain speech), indicative of more pronounced amplitude modulations.

Experiment 2 presented the plain and Lombard sentences in the NiCLS corpus, collected in Experiment 1, mixed with noise to a set of native Dutch listeners. Based on earlier studies reporting an intelligibility benefit of Lombard speech, we predicted Lombard speech to be more intelligible than plain speech, even when matched in overall intensity to plain speech. Moreover, we predicted to find an effect of amplitude modulation power on intelligibility, such that those talkers who produced more pronounced amplitude modulations would also be more intelligible in noise.

Finally, Experiment 3 tested whether the enhanced amplitude modulations in Lombard speech contribute to intelligibility in noise. To that end, the amplitude modulations of Lombard speech were ‘transplanted’ onto matching plain speech sentences. If the resulting ‘transplanted’ speech is more intelligible in noise than the original plain speech, this would suggest that the enhanced amplitude modulations in Lombard speech aid intelligibility in noise.

I. EXPERIMENT 1: speech-in-noise production

A. Methods

1. Participants

Forty-six native Dutch participants (40 females, 6 males; mean age = 22, range = 18-30) were recruited from the Max Planck Institute's participant pool. Participants in all experiments reported in this study gave informed consent as approved by the Ethics Committee of the Social Sciences department of Radboud University (project code: ECSW2014-1003-196). One participant was excluded because she reported, after the experiment, to have hearing impairment in one ear; the remainder reported to have normal hearing. Another three participants were excluded due to technical issues. The data of the remaining 42 participants (37 females, 5 males; mean age = 22, range = 19-30) were included in the analyses reported below.

2. Materials and procedure

Participants were seated behind a table with a computer screen inside a double-walled acoustically isolated booth. A Sennheiser ME64 microphone was fixed on the table in front of the computer screen at approximately 25 cm from the talker and directed towards the participant. Recorded signals were passed to an Alesis Multimix 12 USB amplifier prior to digitalization at 44.1 kHz with a Dell Precision T3400 system using a Creative Sound Blaster X-Fi sound card. Participants wore circum-aural Sennheiser GAME ZERO headphones throughout the experiment, including the quiet condition, to ensure that own-voice masking was held at a constant level.

At the beginning of the speech elicitation experiment, participants were told that they would be asked to read out individual sentences from the folk tale 'The tortoise and the setting sun', both in quiet and in noise. This story consisted of 56 sentences of varying length (for details, see the

prompts.csv file in the NiCLS corpus). The experimenter was seated next to the participant, wearing another pair of Sennheiser GAME ZERO headphones. Participants were instructed to speak clearly to ensure intelligibility for the experimenter, who purportedly heard the same noise as the participant. The experimenter marked the participants' speech for accuracy: only sentence productions without any omissions, additions, or hesitations were marked as 'accurate'. This procedure ensured that participants produced speech with communicative intent, which has been shown to enhance speech adjustments in adverse listening conditions (Garnier et al., 2010). Participants were instructed not to move on their seats while talking so as to avoid unnecessary noise in the recordings.

Stimulus presentation was controlled by Presentation software (v16.5; Neurobehavioral Systems, Albany, CA, USA). Sentences were presented on screen one at a time, controlled by the experimenter. Participants always first produced the sentences (in fixed chronological order) in quiet, and then again in noise. In the speech-in-noise block, speech-shaped noise (SSN) was played diotically through headphones at 85 dB A-weighted SPL (calibrated using a Bruel & Kjaer type 4153 artificial ear and a Bruel & Kjaer type 2260 sound level meter). The SSN was constructed by filtering white noise with the long-term average spectrum of the Dutch VU-sentences (both the male and female talker; Versfeld et al., 2000).

3. Acoustic analysis

Any leading and trailing silences around the sentences were manually removed before analysis. The acoustic analysis involved calculating the modulation spectrum of the Lombard vs. plain sentences, similar to the method in Bosker & Cooke (2018). The analysis was performed separately for each individual talker. First, the overall power of each individual recording (root-mean-square; RMS) was normalized, matching the overall energy of the plain and Lombard speech recordings.

Hence, any potential differences between plain and Lombard speech cannot be attributed to differences in overall energy. Then, all the recordings from one particular talker were concatenated one after another, without inserting any silent intervals in between, separately for the two speech conditions (Lombard vs. plain). The two resulting concatenated signals were filtered by a second-order Butterworth band-pass filter spanning the 500-4000 Hz range (covering the most relevant frequency range for speech intelligibility, while excluding variation in fundamental frequency, considering our diverse talker sample), followed by estimation of the envelope of the filter's output via the Hilbert transform. The envelope signal was then submitted to a Fast Fourier Transform, and the computed amplitude of the various modulation frequency components formed the modulation spectrum of one talker in one particular condition. These modulation spectra were binned into bins of 0.5 Hz for visualization purposes.

B. Results

In total, 5152 recordings were made (46 participants * 56 sentences * 2 speech conditions). After exclusion of four participants, 4704 recordings remained. Recordings that had been evaluated as inaccurate by the experimenter, together with the matching plain or Lombard counterpart recording from that talker, were excluded from analysis (n = 736; i.e., 368 recording pairs of which at least one member had been evaluated as inaccurate). The acoustic analysis described above was performed on the remaining 1984 recording pairs (n = 3968).

Figure 1 displays the average modulation spectra of the plain and Lombard speech across all talkers. The difference between the blue (dark gray) and orange (light gray) lines suggests that there is higher power in the modulation spectrum of Lombard speech (compared to plain speech), especially in the lower frequency range between 1-8 Hz. This was statistically assessed by means

of a Linear Mixed Model (LMM; Baayen et al., 2008) as implemented in the lme4 library (version 1.0.5; Bates et al., 2015) in R (R Development Core Team, 2012). We included a fixed effect of Condition (categorical variable; dummy coding, with plain mapped onto the intercept) as predictor, with Talker entered as random factor with by-talker random slopes for Condition (Barr et al., 2013). Statistical significance was assessed by means of log-likelihood model comparison using the `anova()` function in R, comparing the model with the predictor Condition to a simpler model without that predictor. This LMM revealed a significant effect of Condition ($\beta = 0.319$, $SE = 0.027$, $t = 11.630$; model comparison: $\chi^2(3) = 143.96$, $p < 0.001$), indicating that Lombard speech had significantly higher average power in the modulation spectrum compared to plain speech.

To further investigate which modulation frequency bands drove this effect, we built another LMM that additionally included the predictor Frequency Band (categorical variable; dummy coding, rotating which of four octave bands [1-2, 2-4, 4-8, and 8-15 Hz] was mapped onto the intercept), as well as its interaction with Condition. This extended model was a better fit to the data compared to the original LMM, as assessed by log-likelihood model comparison ($\chi^2(6) = 3744.1$, $p < 0.001$), indicating that the effect of Lombard speech was more pronounced in some bands than others. Rotating which level of the predictor Frequency Band was mapped onto the intercept allowed assessment of the statistical significance of Condition in the various bands, using the Satterthwaite approximation, as implemented in the package lmerTest in R, for degrees of freedom (Luke, 2017). This procedure showed that a significant effect of Condition was observed in all bands ($p < 0.001$), except for the 8-15 Hz band ($p = 0.180$). This suggests that the difference between Lombard and plain speech was primarily driven by the lower modulation frequencies.

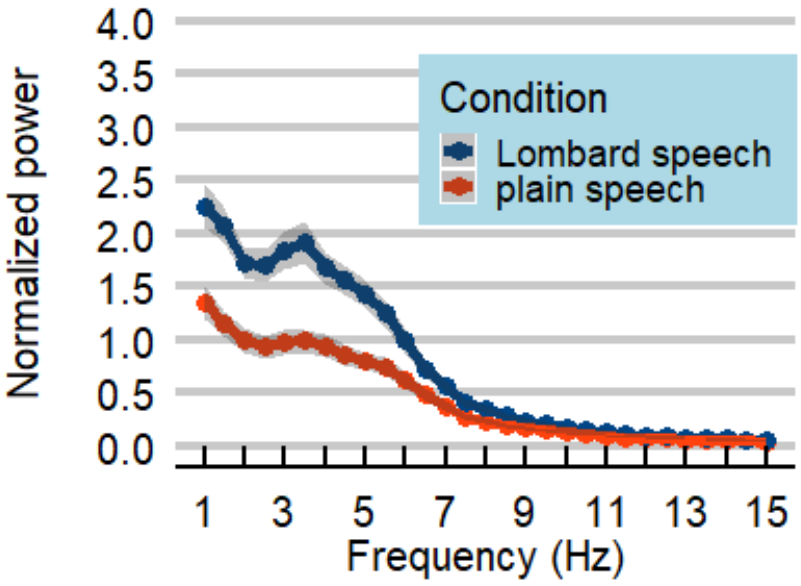

Figure 1. (Color online) **Average modulation spectra of Experiment 1.** Average energy of various modulation frequencies in the Lombard and plain speech of the NiCLS corpus, after normalizing the overall power (RMS) of each recording (hence: "normalized power"). Blue (dark gray) line indicates Lombard speech, orange (light gray) indicates plain speech. Shaded areas enclose 1.96 x SE on either side; that is, the 95% confidence intervals.

C. Interim discussion

The acoustic analysis of the speech produced in Experiment 1 revealed greater power in the modulation spectrum of Lombard speech compared to plain speech. This suggests that the amplitude modulations in Lombard speech were more pronounced. However, the results also suggested that the temporal envelope of Lombard speech is *not* simply an expanded version of the envelope of plain speech. Rather, the effect was primarily driven by the lower frequencies (1-8

Hz). This observation is in line with Bosker and Cooke (2018), who also reported enhanced amplitude modulations in English Lombard speech in the lower frequency range (1-4 Hz). This suggests that the difference in amplitude modulations in Lombard and plain speech may be driven by more pronounced syllabic energy fluctuations in Lombard speech.

II. EXPERIMENT 2: speech-in-noise perception

Having established that Lombard speech has more enhanced amplitude modulations than plain speech in Experiment 1, Experiment 2 set out to assess the contribution of these enhanced amplitude modulations to speech intelligibility. The Lombard and plain speech recordings from Experiment 1 were matched in intensity and presented to listeners in noise. We expected to replicate the well-known intelligibility benefit of Lombard speech, namely that intensity-matched Lombard speech is more intelligible in noise than plain speech. Crucially, if enhanced amplitude modulations contribute to this intelligibility benefit of Lombard speech, we should find that speech with more pronounced amplitude modulations is more intelligible in noise.

A. Methods

1. Participants

Forty-one native Dutch participants (31 females, 10 males; mean age = 23, range = 19-34), that had not participated in Experiment 1, were recruited from the Max Planck Institute's participant pool. Peripheral auditory function was assessed by measuring air-conduction pure-tone thresholds with a PC-based diagnostic audiometer (Oscilla USB-300, Inmedico A/S, Aarhus, Denmark). Pure-tone thresholds were determined at 0.125, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, and 8 kHz in both ears. Five participants with two or more pure-tone thresholds above 20 dB HL were excluded from

analysis. One other participant was excluded due to technical issues. The data of the remaining 35 participants (26 females, 9 males; mean age = 23, range = 19-34) were included.

2. Materials and procedure

The 1984 Lombard and plain speech recording pairs (N = 3968) from Experiment 1 formed the basis of Experiment 2. Stimulus presentation was controlled by Presentation software (v16.5; Neurobehavioral Systems, Albany, CA, USA). Participants were seated behind a table with a computer screen and a keyboard inside the same double-walled acoustically isolated booth as used for Experiment 1. Participants also wore the same circum-aural Sennheiser GAME ZERO headphones as used for Experiment 1. Participants were instructed they would hear spoken sentences from various talkers in loud noise and their task was to type out as many words from the sentences as they could make out.

The spoken sentences from Experiment 1 were matched in overall intensity (‘Scale intensity: 70 dB’ in Praat; Boersma & Weenink, 2016) and presented to participants mixed with speech-shaped noise (SSN). The SSN was constructed by filtering white noise with the long-term average spectrum of all the plain speech in the NiCLS corpus (cf. Cooke, Mayo, et al., 2014). Utterance-plus-noise stimuli were delivered diotically at a -5 dB signal-to-noise ratio (SNR; Cooke et al., 2013), surrounded by noise on- and off-ramps (ramp duration: 500 ms). Following Cooke, Mayo, et al. (2014), the Lombard stimuli and plain stimuli were presented using a blocked design with block presentation order counter-balanced across participants. Each participant heard the plain speech version of a particular sentence produced by a given talker in one block, and the Lombard speech version of that same sentence from that same talker in the other block. Within a block, each participant heard each of the 56 sentences once (but in a unique random order), hearing as many different talkers as possible (given the uneven design in the 1984 selected speech recording pairs).

Participants were allowed to take a short break in between blocks.

B. Results

On average, participants correctly identified 42% and 63% of the words in plain vs. Lombard speech, respectively. This shows a Lombard speech 'intelligibility benefit' of 21 percentage points (p.p.), which – for comparison – is slightly larger than the 18 p.p. in Lu and Cooke (2009) and the 16 p.p. in Cooke, Mayo, et al. (2014). Figure 2 shows the average intelligibility of the plain and the Lombard speech conditions.

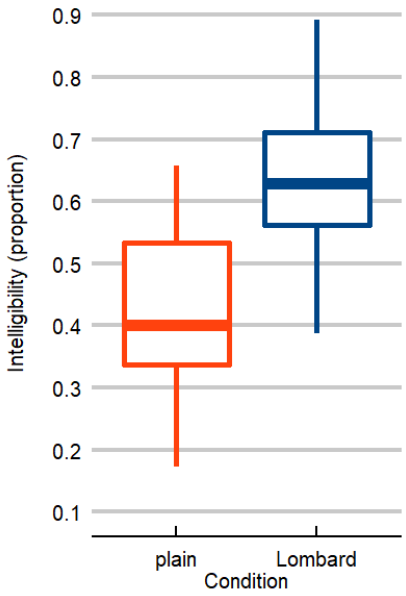

Figure 2. **Intelligibility of plain and Lombard speech.** Intelligibility in proportion words correct in the plain vs. the Lombard speech conditions (presented in SSN at -5 dB SNR).

Proportion correct scores were entered into a Generalized Linear Mixed Model (GLMM; Quené

& Van den Bergh, 2008) with a logistic linking function, as implemented in the lme4 library in R, with weights specified as the maximum number of correct words per sentence. For each talker, the average normalized power in Lombard and plain speech was calculated (larger values for talkers who produced more pronounced amplitude modulations; cf. the x-axis in Figure 3). These values were entered into the GLMM as the predictor Power (numerical variable; using standardized scores to improve model convergence) together with the predictor Condition (categorical variable; dummy coding, with plain mapped onto the intercept). Adding the interaction term to the model did not improve model fit as assessed by log-likelihood model comparison. As random factors, Listener and Sentence were entered as random intercepts with by-listener and by-sentence random slopes for Condition and Power (Barr et al., 2013).

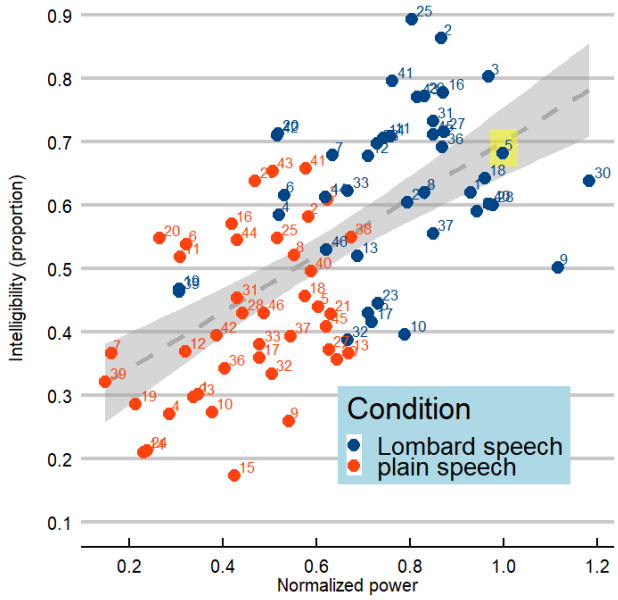

Figure 3. (Color online) **Intelligibility of plain and Lombard speech as a function of average normalized power for individual talkers.** Intelligibility (in proportion words correct) of individual talkers (identified by numbers) in the plain (orange; light gray) vs. the Lombard speech (blue; dark gray) conditions (presented in SSN at -5 dB SNR) as a function of the average normalized power for individual talkers (larger values indicate more pronounced amplitude modulations). The yellow rectangle in the top right corner highlights the data point for Lombard speech produced by Talker 5 (the model talker in Experiment 3). The dashed line shows a fitted logistic function across all data points, with the shaded area enclosing 1.96 x SE on either side; that is, the 95% confidence intervals.

This GLMM revealed a significant effect of Condition (β = 0.689, *SE* = 0.169, z = 4.088, p < 0.001), providing evidence for an overall Lombard speech intelligibility benefit: Lombard speech was more intelligible in noise than plain speech. Additionally, an independent effect of Power was observed (β = 0.289, *SE* = 0.064, z = 4.527, p < 0.001), demonstrating that speech with more pronounced amplitude modulations is more intelligible in noise.

C. Interim discussion

Experiment 2 replicated the well-known intelligibility benefit of Lombard speech, demonstrating that intensity-matched Lombard speech is more intelligible in noise than plain speech. Critically, speech with more pronounced amplitude modulations was found to be more intelligible in noise. This suggests that the more pronounced amplitude modulations in Lombard speech, as observed in Experiment 1, contribute to the intelligibility benefit of Lombard speech in noise.

III. EXPERIMENT 3: transplanting amplitude modulations

Experiment 3 was designed to assess the contribution of enhanced amplitude modulations to speech-in-noise intelligibility by means of acoustic manipulations. Results from Experiment 1 showed that the more pronounced character of the amplitude modulations in Lombard speech was not a matter of linear scaling: rather, the effect varied across different modulation frequencies. Therefore, we decided against using a linear expansion technique and instead opted for prosody transplantation. Experiment 3 involved another speech-in-noise listening experiment. Participants were presented with three speech conditions: the original plain speech, the original Lombard speech, and ‘transplanted speech’. This ‘transplanted speech’ was constructed by transplanting the amplitude modulations from Lombard speech onto the plain speech recordings. If more pronounced amplitude modulations contribute to the Lombard speech intelligibility benefit, we should find that ‘transplanted speech’ is more intelligible in noise than the original plain speech.

A. Methods

1. Participants

42 native Dutch participants (32 females, 10 males; mean age = 22, range = 19-29) , that had not participated in Experiments 1-2, were recruited from the Max Planck Institute's participant pool. Peripheral auditory function was assessed using the same pure-tone threshold assessment as in Experiment 2. Six participants with two or more pure-tone thresholds above 20 dB HL were excluded from analysis. The data of the remaining 26 females and 10 males (mean age = 22, range = 19-29) were included.

2. Materials and procedure

The plain speech materials from Experiment 1-2 were manipulated to have the same intensity contour as Lombard speech in the following fashion. First, talker 5 (highlighted in Figure 3) was selected as the model talker, because (1) this talker produced very pronounced amplitude modulations in Lombard speech; and (2) only 2 out of the 56 sentences from this talker were excluded in Experiment 1, meaning that 54 sentences were available as model sentences. In total, there were 1923 plain speech sentences that could be matched to the Lombard speech of talker 5.

Each plain speech recording was paired to the matching Lombard recording from talker 5. After matched in overall intensity, the temporal characteristics of the Lombard speech were dynamically time warped (DTW) to match those of the plain speech (mostly involving compression, considering that Lombard speech is typically slower than plain speech). Following Cooke, Mayo, et al. (2014), we used a combination of dynamic time warping and PSOLA techniques as implemented in the Revoice Pro 3 program (Synchroarts), as illustrated in Figure 4. This process ensures that the phonetic content of the plain and Lombard speech is aligned in time, which forms a prerequisite for transplanting the Lombard intensity contour onto the corresponding plain speech

signal. Then, in Praat (Boersma & Weenink, 2016), the plain signal was multiplied by its own inverse intensity contour, after which it was multiplied by the intensity contour of the DTW Lombard speech. This resulted in *transplanted speech* that was identical to the original plain speech, except that it contained the intensity contour of the (DTW) Lombard speech from talker 5 (intensity contours of middle and bottom signal in Figure 4 are identical).

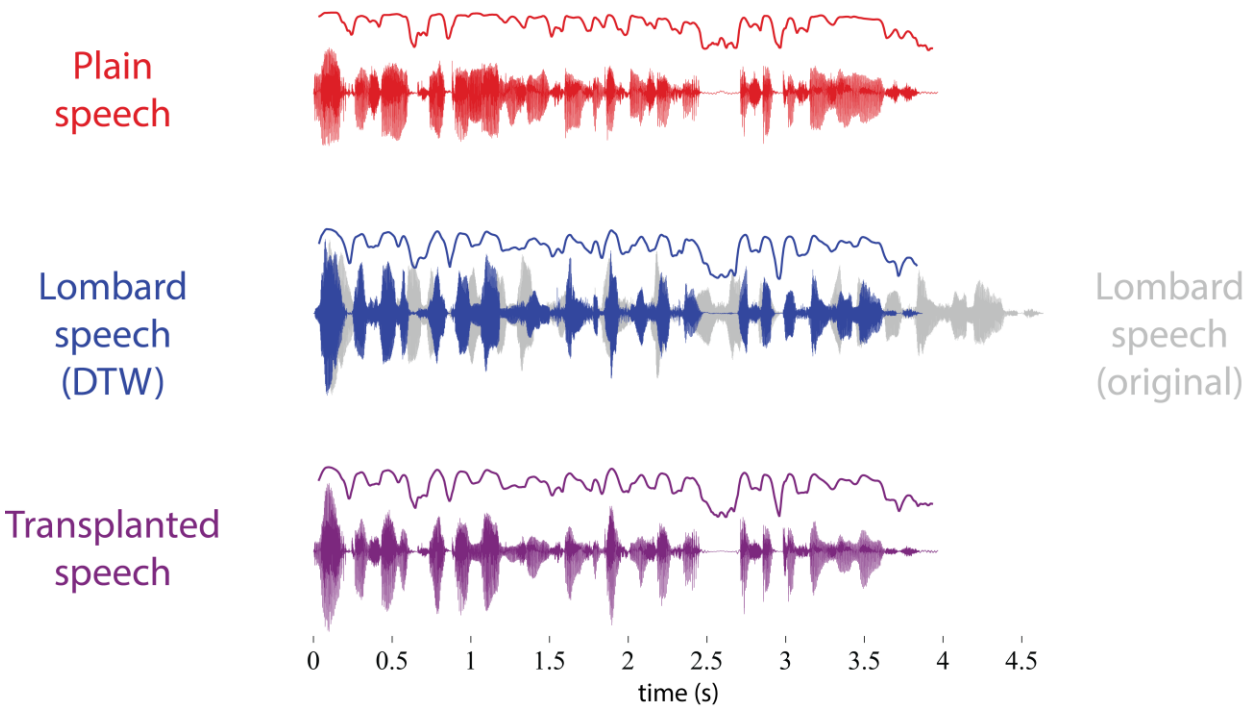

Figure 4. (Color online) **Example of transplantation method.** The top signal (red) shows an example plain speech sentence (sentence 19) from talker 1. The hidden middle signal (light gray) shows the matching Lombard speech sentence from talker 5 (the model talker), which has a longer duration than the plain speech (i.e., Lombard speech is slower than plain speech). This Lombard speech signal was first dynamically compressed (dynamic time warping; DTW) to match the temporal dynamics of the plain speech, resulting in the middle signal in blue. Finally, the intensity contour (individual lines above wave forms) of this signal

was transplanted onto the plain speech, resulting in the bottom signal (purple). This transplanted speech is identical to the plain speech except for more pronounced amplitude modulations.

Participants in Experiment 3 were tested using the same lab, devices, software, and instructions as in Experiment 2. Speech stimuli (plain, transplanted, and Lombard) were matched in overall intensity and presented to participants together with the same SSN as in Experiment 2. However, in Experiment 3, utterance-plus-noise stimuli (plus noise on- and off-ramps of 500 ms) were delivered at an SNR of -3 dB (instead of -5 dB SNR in Experiment 2). Note that our primary interest was in the comparison of plain vs. transplanted speech. Previously, in Experiment 2, the average intelligibility of plain speech at an SNR of -5 dB was 42%. Increasing the SNR by 2 dB in Experiment 3 would enhance overall intelligibility, thus boosting participants' motivation in the (difficult) transcription task.

The plain, transplanted, and Lombard stimuli were presented using a blocked design with block presentation order counter-balanced across participants. Each participant heard the plain speech version of a particular sentence produced by a given talker in one block, and the transplanted and Lombard speech versions of that same sentence from that same talker in the other two blocks. Within a block, each participant heard each of the 56 sentences once, hearing as many different talkers as possible (given the uneven design in the 1923 selected speech recordings). Participants were allowed to take a short break in between blocks.

B. Results

On average, participants correctly identified 51% and 82% of the words in plain and Lombard speech conditions, respectively. Interestingly, participants correctly identified more words in the

transplanted speech (68% of the words) vs. plain speech condition (51%; see Figure 5).

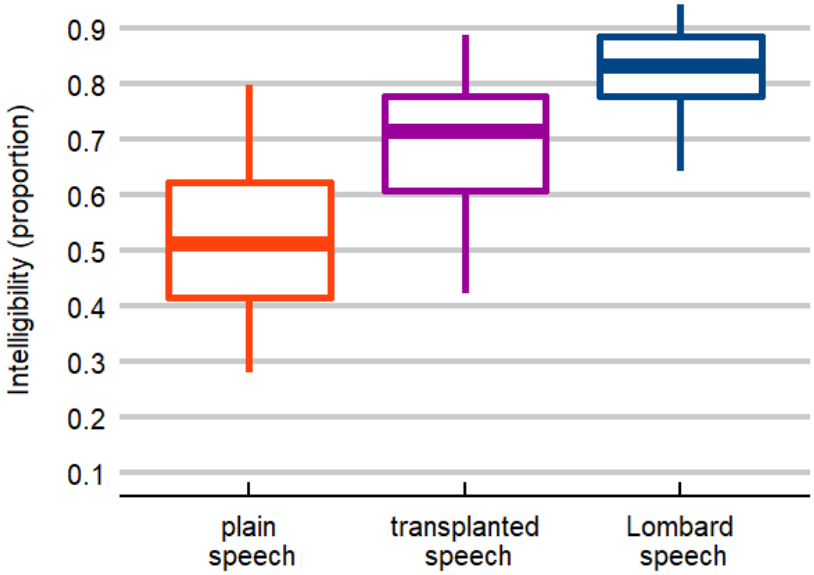

Figure 5. (Color online) **Intelligibility of plain, transplanted, and Lombard speech in Experiment 3.** Speech materials were presented in SSN at -3 dB SNR.

In order to statistically test this difference, we entered the proportion correct scores into a GLMM with a logistic linking function, as implemented in the lme4 library in R, with weights specified as the maximum number of correct words per sentence. This GLMM included the predictor Condition (categorical variable; dummy coded, with plain mapped onto the intercept), including two contrasts: comparing the intelligibility between plain vs. transplanted and plain vs. Lombard. Listener and Sentence were entered as random intercepts with by-listener and by-sentence random slopes for Condition and Power.

This GLMM revealed that the Lombard speech was more intelligible than the plain speech (β = 1.734, SE = 0.144, z = 12.019, p < 0.001). More critically, the GLMM also established that the

transplanted speech was more intelligible than the plain speech (β = 0.805, *SE* = 0.129, z = 6.253, $p < 0.001$). Mapping transplanted speech onto the intercept of the predictor Condition revealed that the difference between transplanted speech and Lombard speech was also statistically significant (β = 0.929, *SE* = 0.117, z = 7.936, $p < 0.001$).

Since part of the intelligibility benefit of Lombard speech appears to originate in energetic masking release (Lu & Cooke, 2009), a glimpsing analysis (Cooke, 2006) was performed to estimate the proportion of time-frequency regions where the target speech was likely to be audible. Individual sentences and their corresponding masker waveforms were separately processed through a 55 channel gammatone filterbank with center frequencies ranging from 100 to 8000 Hz on an ERB-rate scale. A time-frequency representation was constructed by extracting the Hilbert envelope at the output of each filter followed by smoothing with a leaky integrator with an 8 ms time constant and downsampling to 100 Hz. Glimpse proportions, defined as the proportion of time-frequency cells in this representation where the speech energy exceeded that of the masker, were subjected to a new GLMM, very similar in structure to the GLMM above used for the intelligibility scores. This GLMM estimated the glimpse proportions as a function of the predictor Condition (same coding as above) with a logistic linking function, with weights specified as the maximum number of time-frequency cells per sentence. Talker and Sentence were entered as random intercepts with by-talker and by-sentence random slopes for Condition. In line with Lu and Cooke (2009), Lombard speech had a substantially higher glimpse proportion than plain speech (0.128 vs. 0.093; β = 0.363, *SE* = 0.026, z = 14.020, $p < .001$). However, the glimpse proportion of 0.092 for transplanted speech was almost identical to that of plain speech (β = -0.012, *SE* = 0.012, z = -1.063, p = 0.289).

Moreover, adding the (scaled) glimpsing proportions as a predictor to the GLMM analyzing the

intelligibility data from Experiment 3 revealed that (i) greater glimpsing proportions indeed correlated with intelligibility (β = 0.572, *SE* = 0.015, z = 38.411, $p < 0.001$); yet (ii) adding this predictor to the model did not qualitatively change the Condition effects. This demonstrated that the Condition effects were observable even when effects of audibility were partialled out.

C. Interim discussion

Experiment 3 replicated Experiment 2 by once more revealing a Lombard speech 'intelligibility benefit' in noise. Crucially, it also showed that the transplanted speech was more intelligible than the plain speech. This result suggests that the enhanced amplitude modulations in Lombard speech contribute to the intelligibility benefit: when plain speech is manipulated to have the same intensity contour as Lombard speech, intelligibility increases. The glimpsing analysis suggested that the intelligibility benefit of transplanted speech was not due to energetic masking release.

IV. GENERAL DISCUSSION

This study addressed the question what makes speech produced in noise more intelligible in noise compared to speech produced in quiet: the Lombard speech intelligibility benefit. We specifically targeted the contribution of *enhanced amplitude modulations* in the temporal envelope of Lombard speech. This aspect of Lombard speech has received relatively little attention, while there are clear indications in the literature that speech intelligibility greatly relies on these amplitude modulations (Drullman et al., 1994a; Elliott & Theunissen, 2009; Shannon et al., 1995; Smith et al., 2002).

Experiment 1 introduced the first Dutch corpus of Lombard speech (NiCLS). Acoustic analysis of plain vs. Lombard speech recordings, with matched overall intensity, revealed greater power in

the modulation spectrum of Lombard speech compared to plain speech, particularly in the lower frequency range (1-8 Hz). This suggests that the amplitude modulations in Lombard speech were more pronounced, particularly involving energy fluctuations at (roughly) the syllabic rate, extending earlier observations in English (Bosker & Cooke, 2018) to a new language: Dutch. Because the same effect has been found across two different languages, across different corpora with different elicitation techniques, different sentence materials, and different noise types (cf. Saigusa & Hazan, 2019), this effect is likely to be robust and may generalize to everyday spoken communication.

Experiment 2 involved a perception experiment, assessing the intelligibility of the Lombard and plain sentences in the NiCLS corpus with matched overall intensity when presented in noise. Proportion word correct scores were higher for Lombard speech compared to plain speech, supporting the Lombard speech intelligibility benefit in noise. More interestingly, individual talkers' overall intelligibility correlated with the normalized power of amplitude modulations in their speech. That is, those talkers who produced more pronounced amplitude modulations were also more intelligible in noise.

This observation corroborates the central role that amplitude modulations play in speech perception (Flinker et al., 2019; Ghitza, 2012; Shannon et al., 1995; Smith et al., 2002). Moreover, it reveals the contribution of amplitude modulations to speech intelligibility on an individual-talker level. This finding extends to other studies looking into the acoustic correlates of speaker intelligibility (e.g., Bradlow et al., 1996), with implications for speech synthesis and speech recognition strategies, and for special populations (e.g., hearing-impaired; non-natives) who are particularly sensitive to intelligibility differences among talkers.

Although Experiment 2 suggests that enhanced amplitude modulations in Lombard speech

improve intelligibility in noise, this evidence is correlational. Therefore, Experiment 3 manipulated the amplitude modulations in plain speech by means of prosody transplantation. We constructed 'transplanted' speech by transplanting the amplitude modulations from Lombard speech onto the plain speech recordings. Participants in Experiment 3 listened to (intensity-matched) plain speech, Lombard speech, and 'transplanted' speech in speech-shaped noise, this time at an SNR of -3 dB. Results showed, again, that Lombard speech was more intelligible than plain speech. More critically, participants scored higher proportion word correct scores for 'transplanted' speech compared to the original 'plain' speech, suggesting a link between the power of amplitude modulations in the temporal envelope speech and speech intelligibility. Hence, this suggests that the enhanced amplitude modulations present in Lombard speech contribute to the intelligibility benefit of Lombard speech in noise.

Note, however, that our transplantation technique – beyond transplanting the enhanced nature of the amplitude modulations in Lombard speech – may have transferred other characteristics of the amplitude modulations in Lombard speech as well. For instance, if the modulation energy in Lombard speech is not only more pronounced but also differently allocated across the utterance, then these two characteristics are correlated within the design of Experiment 3. The fact that the modulation power difference between Lombard and plain speech was mainly driven by the lower frequencies (1-8 Hz) indeed suggests that the envelope of Lombard speech is not simply an expanded version of the envelope of plain speech, which motivated us to opt for the transplantation technique (i.e., not for simply expanding the envelope of plain speech). As a result, we do not claim that the enhanced modulation power in Lombard speech is the *only* factor that drives the difference between plain and transplanted speech in Experiment 3. Nevertheless, the outcomes of Experiment 3 do demonstrate that the temporal envelope of Lombard speech contains critical information for

its intelligibility benefit. It was the only property that was altered by the transplantation technique. Earlier studies primarily found intelligibility effects of spectral manipulations (e.g., Cooke, Mayo, et al., 2014; Godoy et al., 2014; Lu & Cooke, 2009). This present finding builds on, yet goes beyond previous literature that artificially manipulated the modulation spectrum of speech in an attempt to improve intelligibility. First, we show intelligibility improvements when plain speech was manipulated to have more pronounced amplitude modulations. This is in contrast to some earlier studies that failed to find intelligibility improvements, or even reported intelligibility decrements after artificially increasing the modulation depth in the temporal envelope (Krause & Braida, 2009; Kusumoto et al., 2005). Second, while our 'transplanted' speech contained the intensity contour as taken from naturally occurring Lombard speech, other researchers manipulated the amplitude modulation components in the speech beyond what is observed even for clear speech (Krause & Braida, 2009). Thus, these results carry implications for our understanding of Lombard speech as occurring in natural communicative situations. They highlight the importance of speech enhancement techniques that are guided by naturally occurring speech, as for instance reported in Koutsogiannaki and Stylianou (2016) for clear speech.

The observed difference between transplanted and plain speech in Experiment 3 raises the question which perceptual and neurobiological mechanisms underlie the beneficial effect of enhanced amplitude modulations on intelligibility in noise. It could be argued that enhanced amplitude modulations would make the target speech 'rise above the noise', producing greater energetic masking release. However, the glimpsing analysis in Experiment 3 did not reveal a higher proportion of time-frequency regions where the target speech was likely to be audible in transplanted vs. plain speech. Therefore, the greater intelligibility of transplanted (compared to plain) speech is unlikely to be accounted for by differences in energetic masking release.

Instead, we interpret the outcomes of the present study in light of neurobiological models of speech perception (Ghitza, 2011; Giraud & Poeppel, 2012; Peelle & Davis, 2012) that posit a central role for endogenous theta oscillations closely following the syllabic rhythm of speech (Arnal et al., 2015; Bosker & Ghitza, 2018; Bosker & Kösem, 2017; Kösem et al., 2018). Applying these models to speech-in-noise and speech-in-speech comprehension, a range of electrophysiological studies have provided evidence that listeners' envelope-tracking response to an attended speaker is amplified compared to an ignored speaker (Dai et al., 2018; Ding & Simon, 2012; Golumbic, Cogan, et al., 2013; Lakatos et al., 2008; Mesgarani & Chang, 2012). This dynamic neural representation of the temporal structure of the attended speech stream (e.g., in a noisy environment, or with a competing speech signal) is thought to function as an amplifier and a temporal filter, aiding speech comprehension in challenging listening conditions. Clearly, the outcomes of the present behavioral study do not give a definitive answer on the debate about the role of neural oscillations in speech comprehension. Still, arguing from these oscillatory frameworks, we speculate that the enhanced amplitude modulations in Lombard speech (and hence also in the 'transplanted' speech in Experiment 3) help the listening brain to 'track' the attended talker, aligning neuronal excitability to the temporal structure of the attended signal, thus facilitating speech-in-noise perception. Future neuroimaging studies could investigate the neurobiological mechanisms underlying the Lombard speech intelligibility benefit, for instance by assessing whether the more pronounced amplitude modulations in Lombard speech indeed facilitate cortical speech-tracking, aiding speech-in-noise intelligibility.

V. DATA AVAILABILITY STATEMENT

The NiCLS corpus is available for download from: https://hdl.handle.net/1839/21ee5744-b5dc-

VI. ACKNOWLEDGEMENTS

This research was supported by the Max Planck Society for the Advancement of Science, Munich, Germany (H.R.B.) and by funding from the Basque Government Consolidados grant to the Language and Speech Laboratory at the University of the Basque Country (M.C.). We would like to thank Chris-Jan Beerendonk for technical support, Esther Janse and Margret van Beuningen from the Centre for Language Studies, Radboud University, Nijmegen, for providing the audiometer, Rosemarije Weterings, Esther de Kerf, and Inge Pasman for their help in testing participants, the student-assistants of the Psychology of Language department of the Max Planck Institute for Psycholinguistics for their help in coding participants' responses, and Annelies van Wijngaarden who coordinated their efforts.

VII. REFERENCES

Arnal, L. H., Giraud, A.-L., & Poeppel, D. (2015). A Neurophysiological Perspective on Speech Processing in "The Neurobiology of Language." In G. Hickok & S. Small (Eds.), *Neurobiology of Language* (pp. 463–478). Academic Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/doi:10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer [computer program]*.

Bosker, H. R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. *Attention,*

Perception & Psychophysics, *79*, 333–343. https://doi.org/10.3758/s13414-016-1206-4

Bosker, H. R., & Cooke, M. (2018). Talkers produce more pronounced amplitude modulations when speaking in noise. *Journal of the Acoustical Society of America*, *143*. https://doi.org/10.1121/1.5024404

Bosker, H. R., & Ghitza, O. (2018). Entrained theta oscillations guide perception of subsequent speech: behavioural evidence from rate normalisation. *Language, Cognition and Neuroscience*, *33*(8), 955–967. https://doi.org/10.1080/23273798.2018.1439179

Bosker, H. R., & Kösem, A. (2017). An entrained rhythm's frequency, not phase, influences temporal sampling of speech. In *Proceedings of Interspeech 2017, Stockholm*. dx.doi.org/10.21437/Interspeech.2017-73

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

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*, 1562–1573.

Cooke, M., King, S., Garnier, M., & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech & Language*, *28*, 543–571.

Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., & Tang, Y. (2013). Evaluating the intelligibility benefit of speech modifications in known noise conditions. *Speech Communication*, *55*, 572–585.

Cooke, M., Mayo, C., & Villegas, J. (2014). The contribution of durational and spectral changes to the Lombard speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *135*, 874–883.

Dai, B., Chen, C., Long, Y., Zheng, L., Zhao, H., Bai, X., Liu, W., Zhang, Y., Liu, L., Guo, T., Ding, G., & Lu, C. (2018). Neural mechanisms for selectively tuning in to the target speaker in a naturalistic noisy situation. *Nature Communications*, *9*(1), 2405. https://doi.org/10.1038/s41467-018-04819-z

Ding, N., Patel, A., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience and Biobehavioral Reviews*, *81B*, 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, *109*(29), 11854–11859. https://doi.org/10.1073/pnas.1205381109

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage, 85*, 761–768.

Dreher, J. J., & O'Neill, J. (1957). Effects of ambient noise on speaker intelligibility for words and phrases. *The Journal of the Acoustical Society of America, 29*, 1320–1323.

Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech recognition. *Journal of the Acoustical Society of America, 95*, 2670–2680.

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America, 95*, 1053–1064.

Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Computational Biology, 5*, e1000302.

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., & Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nature Human Behaviour, 1*, 393–405. https://doi.org/10.1038/s41562-019-0548-z

Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research, 53*, 588–608.

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology, 2*.

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology, 3*, 238.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience, 15*, 511–517.

Godoy, E., Koutsogiannaki, M., & Stylianou, Y. (2014). Approaching speech intelligibility enhancement with inspiration from Lombard and Clear speaking styles. *Computer Speech & Language, 28*(2), 629–647. https://doi.org/10.1016/j.csl.2013.09.007

Golumbic, E. M. Z., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party." *The Journal of Neuroscience, 33*, 1417–1426.

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson,

R., Mehta, A. D., & Simon, J. Z. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party.” *Neuron*, *77*, 980–991.

Golumbic, E. M. Z., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language*, *122*, 151–161.

Hotchkin, C., & Parks, S. (2013). The Lombard effect and other noise-induced vocal modifications: insight from mammalian communication systems. *Biological Reviews*, *88*, 809–824.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional Gain Control of Ongoing Cortical Speech Representations in a “Cocktail Party.” *Journal of Neuroscience*, *30*(2), 620–628. https://doi.org/10.1523/JNEUROSCI.3631-09.2010

Kösem, A., Bosker, H. R., Takashima, A., Jensen, O., Meyer, A., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, *28*(18), 2867–2875. https://doi.org/10.1016/j.cub.2018.07.023

Koutsogiannaki, M., & Stylianou, Y. (2016). Modulation Enhancement of Temporal Envelopes for Increasing Speech Intelligibility in Noise. In *Proceedings of Interspeech* (pp. 2508–2512).

Krause, J. C., & Braida, L. D. (2009). Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech. *The Journal of the Acoustical Society of America*, *125*, 3346–3357.

Kusumoto, A., Arai, T., Kinoshita, K., Hodoshima, N., & Vaughan, N. (2005). Modulation enhancement of speech by a pre-processing algorithm for improving intelligibility in reverberant environments. *Speech Communication*, *45*(2), 101–113. https://doi.org/10.1016/j.specom.2004.06.003

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*, 110–113.

Lombard, E. (1911). Le signe de l’elevation de la voix [The sign of the rise in the voice]. *Annales Des Maladies de l’Oreille et Du Larynx [Annals of Ear and Larynx Diseases]*, *37*, 101–119.

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

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4),

1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Luo, J., Goerlitz, H. R., Brumm, H., & Wiegrebe, L. (2015). Linking the sender to the receiver: vocal adjustments by bats to maintain signal detection in noise. *Scientific Reports*, *5*, 18556. https://doi.org/10.1038/srep18556

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*, 233–236.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00320

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

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413–425.

R Development Core Team. (2012). *R: A Language and Environment for Statistical Computing [computer program]*.

Rimmele, J. M., Golumbic, E. M. Z., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex*, *68*, 144–154.

Saigusa, J., & Hazan, V. (2019). The effect of temporally fluctuating maskers on speech production and communication. In *Proceedings of the 19th International Congress of Phonetic Sciences 2019 [ICPhS XIX], Melbourne* (p. 5).

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*, 87–90.

Steeneken, H. J., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America*, *67*, 318–326.

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

928.

Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., & Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America*, *142*, 1976–1989.

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *The Journal of the Acoustical Society of America*, *107*, 1671–1684.

VIII. FIGURE CAPTIONS

Figure 1. (Color online) **Average modulation spectra of Experiment 1.** Average energy of various modulation frequencies in the Lombard and plain speech of the NiCLS corpus, after normalizing the overall power (RMS) of each recording (hence: "normalized power"). Blue (dark gray) line indicates Lombard speech, orange (light gray) indicates plain speech. Shaded areas enclose 1.96 x SE on either side; that is, the 95% confidence intervals.....12

Figure 2. **Intelligibility of plain and Lombard speech.** Intelligibility in proportion words correct in the plain vs. the Lombard speech conditions (presented in SSN at -5 dB SNR). 15

Figure 3. (Color online) **Intelligibility of plain and Lombard speech as a function of average normalized power for individual talkers.** Intelligibility (in proportion words correct) of individual talkers (identified by numbers) in the plain (orange; light gray) vs. the Lombard speech (blue; dark gray) conditions (presented in SSN at -5 dB SNR) as a function of the average normalized power for individual talkers (larger values indicate more pronounced amplitude modulations). The yellow rectangle in the top right corner highlights the data point for Lombard speech produced by Talker 5 (the model talker in Experiment 3). The dashed line shows a fitted logistic function across all data points, with the shaded area enclosing 1.96 x SE on either side; that is, the 95% confidence intervals. 17

Figure 4. (Color online) **Example of transplantation method.** The top signal (red) shows an example plain speech sentence (sentence 19) from talker 1. The hidden middle signal (light gray) shows the matching Lombard speech sentence from talker 5 (the model talker), which has a longer duration than the plain speech (i.e., Lombard speech is slower than plain speech). This Lombard speech signal was first dynamically compressed (dynamic time warping; DTW) to match the temporal dynamics of the plain speech, resulting in the middle signal in blue. Finally, the intensity contour (individual lines above wave forms) of this signal was transplanted onto the plain speech, resulting in the bottom signal (purple). This transplanted speech is identical to the plain speech except for more pronounced amplitude modulations. 20

Figure 5. (Color online) **Intelligibility of plain, transplanted, and Lombard speech in Experiment 3.** Speech materials were presented in SSN at -3 dB SNR.....22