THE LOMBARD EFFECT IN MRI NOISE

Amelia J. Gully, Paul Foulkes, Peter French, Philip Harrison, Vincent Hughes

Department of Language and Linguistic Science, University of York, UK
amelia.gully | paul.foulkes | peter.french | philip.harrison | vincent.hughes@york.ac.uk

ABSTRACT

In recent years, magnetic resonance imaging (MRI) has been used extensively to study the vocal tract. MRI equipment makes a loud continuous noise throughout the imaging process. The acoustic properties of speech are known to change in noisy conditions - a phenomenon known as the Lombard effect. The characteristics of Lombard speech have been shown to vary with the type of noise, but no study has yet considered the effect of MRI noise, with its specific spectral properties, on speech. We present results showing that formant values, particularly those for the first formant, vowel space dispersion, and spectral tilt of speech produced in MRI noise is significantly different from speech produced in normal conditions. The effects are both subject- and phoneme-dependent. The results have important implications for all acoustic studies based on MRI data of the vocal tract, and we close with recommendations for collecting such data.

Keywords: Lombard effect, vocal tract modelling, magnetic resonance imaging.

1. INTRODUCTION

When speaking in noise, the human voice changes in a number of ways. Most commonly, amplitude and pitch increase, and spectral tilt decreases [18]. Alterations in formant frequencies have also been reported [14], as have differences in articulation of the tongue, jaw and lips [19]. The combination of these changes is known as the Lombard effect [13]. The degree of change of each parameter depends upon the amplitude and spectral content of the masking noise [17], with most existing studies considering broadband and/or speech-like noise [5].

In the last few decades, magnetic resonance imaging (MRI) has been used extensively for research on speech and the vocal tract (see [16] for a review). Speech produced in an MRI scanner is unlike normal speech in a number of ways. Most scanners require a supine rather than an upright posture, and effect of position on the voice has been considered in a number of studies, reviewed in [16]. In order to capture 3D images, it is usually necessary for the subject to hold a single vocal tract position for an unnaturally long time, leading to hyperarticulation [7]. The MRI scanner is also a very noisy environment, which is likely to elicit the Lombard effect, but to our knowledge there has been no systematic study of the Lombard effect for speech produced in MRI noise. It is therefore essential to determine the characteristics of MRI-specific Lombard speech, in order to properly interpret MRI data of the vocal tract, and audio recordings that may have been captured simultaneously. Additionally, due to the noise and magnetic field in the MRI scanner, corresponding audio recordings are often made separately in a quiet environment. MRI noise may not be fully reproduced during this stage. Since MRI data is often used to inform acoustic models of the vocal tract (e.g. [10, 2]), it is important that comparisons based on such recordings are valid.

MRI noise is generated by the periodic vibrations of large electromagnetic coils, which are amplified by resonances in the structures of the machine and accompanied by the noise of cooling pumps [4], leading to noise levels exceeding 100dB-SPL at the subject's ear [6]; such noise levels are well above the minimum threshold for inducing the Lombard effect [12]. The Lombard effect is known to differ depending on the spectral and temporal characteristics of the masking noise [17]. It is therefore important to study speech produced in MRI noise, which is periodic and thus has different spectral content than the broadband or speech-like noise investigated in previous studies. The size of the Lombard effect has been shown to differ significantly between individual speakers [11]. Further, the communication element of the speech task must be considered as the magnitude of the Lombard effect is greater in communicative than non-communicative tasks [9].

This paper presents a comparison of speech produced in the presence of MRI noise with that produced in quiet, for both standing and supine positions. The remainder of the paper is laid out as follows. Section 2 presents the methods used to collect and analyze the speech data, Section 3 describes and discusses the results, before concluding and offering recommendations for MRI data capture.

2. METHOD

2.1. Data collection

The dataset comprises recordings of speech from 16 British native English speakers, 8 male and 8 female, of which 4 males and 4 females were 'experts' (defined as having a PhD or equivalent experience in phonetics or voice acoustics), and 4 males and 4 females were 'non-experts'. Subjects were split into blocks of four participants, with each block representing a different combination of sex and expertise.

Each subject was recorded in four conditions: standing in noise (StN), supine in noise (SuN), standing in quiet (StQ), and supine in quiet (SuQ). Condition order varied systematically within blocks. The MRI noise used was recorded using a Sennheiser MO 2000 optical microphone, in a GE 3T Signa Excite MRI scanner, during a scan protocol developed for 3D volumetric imaging of the vocal tract (head neurovascular array coil, parallel imaging factor 2, 3D GRE sequence, TR 4.736ms, TE 1.68ms, FA 5°, FOV 384mm with 192x192 acquisition matrix, interpolated to 512x512; 80 contiguous 2mm sagittal slices with no gap). As it is not possible to record the sound pressure level inside the scanner during imaging, due to the magnetic field, noise was played back at a volume that completely blocked auditory feedback of their own voice, replicating MRI conditions. Noise was presented over in-ear headphones covered by ear defenders, which were both removed during the quiet conditions.

Three passages were read in each condition: the North Wind and the Sun, the Grandfather passage, and the Rainbow passage. Since these passages contain unusual words and phrasing, participants were sent the texts in advance and asked to practice reading them aloud. Five held vowels [i, a, ɑ, ɔ, u] were also captured, described to participants as the vowels in 'fleece', 'trap', 'car', 'thought', and 'goose' respectively. Since MRI noise has a pitch, and subjects often unintentionally match this pitch while holding vowels, they were instructed to do so, and the same pitch was provided to participants before held vowel production in the quiet conditions, following [1].

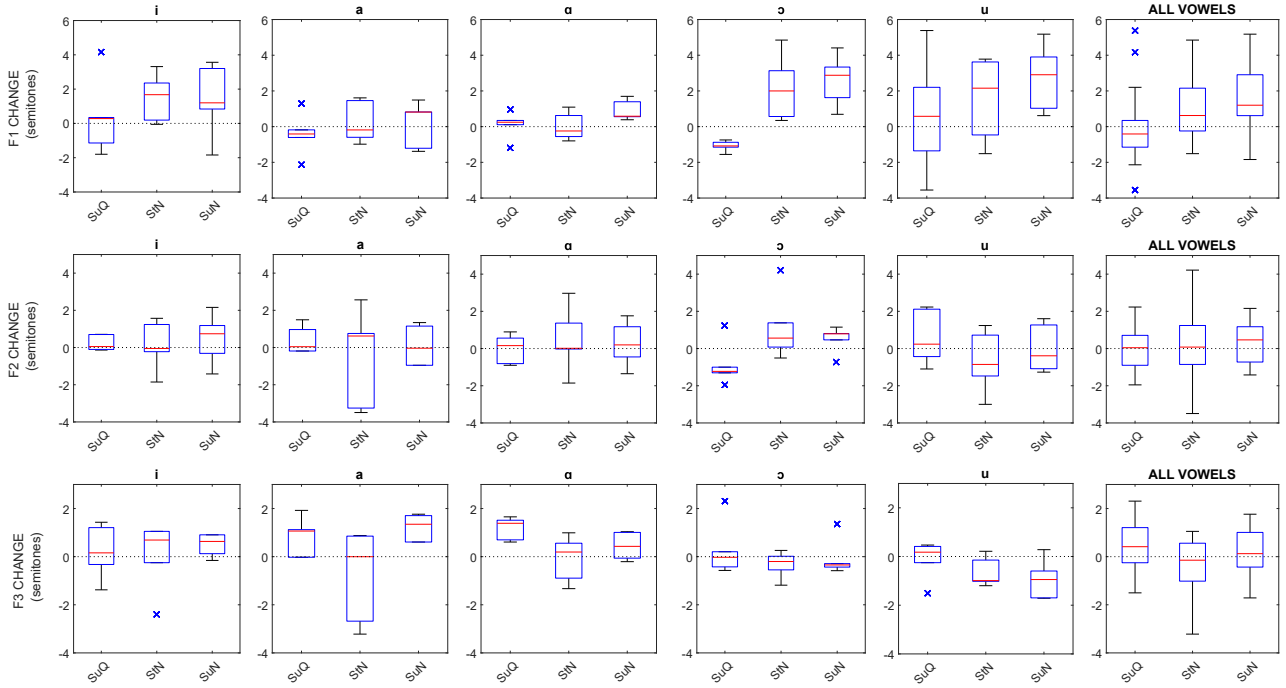
The data were recorded in an anechoic chamber using a Zoom F8 recorder with 48kHz sample rate, and include recordings from: a head-mounted microphone (DPA 4066) taped into place approximately 2.5cm from the corner of the subject's mouth; a measurement microphone (Earthworks M30) placed 50cm away from the mouth (on-axis); an electrolaryngograph to capture vocal fold behaviour; and a speakerphone (Samsung Galaxy S7) mounted approximately 70cm away, recorded via telephone intercept, to provide an additional lower-quality audio channel for future research on robustness of speech measures to channel degradation. The phone also provides a convenient method of establishing consistency of communicative intent, which is known to affect the size of the Lombard effect [9]: in all conditions, participants were asked to read the passages so that a listener on the other end of the phone line could hear them clearly. Participants could see the phone during all conditions but received no feedback from it. In a real MRI study, participants would not see a phone; therefore if separate audio recordings are to be made, different but consistent instructions should be given about communicative intent in both environments. Note also that using a phone may introduce some Lombard effect even in quiet conditions [12].

2.2. Data Analysis

In the following, only audio from the head-mounted microphone was used. Audio files were manually segmented in Praat [3] to extract five instances each of the vowels [i, a, ɑ, ɔ, u] from running speech, selected to include instances from near the start, middle and end of the reading task, and one second each from the same held vowels. These were used to automatically extract formant values F1–F3 using the formant tracking function in Praat, with sex- and vowel-specific reference values based on [15], using LPC order 12, with no manual correction and the same settings across all vowels. Passages were edited to remove repetitions, coughs etc., and these corrected passages were used to calculate the long term average spectrum (LTAS) of the utterances.

A number of additional properties of Lombard speech are not considered here, since the focus of this study is on the spectral properties of MRI-induced Lombard speech. Pitch and duration changes in particular are characteristic of Lombard speech. The average pitch of running speech was found to increase in MRI noise in this study, by 3.2Hz or 2.77%, but MRI noise in particular may confound these effects as speakers may 'tune' their speech to the pitch of the noise (this is expected for held vowels, as noted above, but the effect on running speech is unknown). Duration changes may also be affected by the generally unpleasant nature of the MRI noise, with subjects reporting either a desire to get the recording over with as quickly as possible, or that the noise was distracting and caused them to speak slowly. These effects and others, such as amplitude and voice quality, will be considered in a forthcoming paper.

Figure 1: Shifts in vowel formants, in semitones, compared to the neutral (standing, quiet) condition. Results are for F1 (top row), F2 (middle row), and F3 (bottom row). Each column shows results for a different vowel with the combined result across all 5 vowels in the rightmost column.



3. RESULTS AND DISCUSSION

This section considers how the presence of MRI noise impacts the spectrum of the resulting speech. Figure 1 illustrates the formant shifts in each condition SuQ, StN, and SuN, compared to the 'neutral' condition StQ. Differences are calculated in semitones to provide an approximate perceptual weighting and permit comparison between male and female subjects. To determine significance, a linear mixed effects model (LMEM) was used with noise condition, position, subject sex, subject expertise, and vowel as fixed effects; subject, and word from which the vowel was taken, were set as random effects. Significance was determined by comparing the full LMEM to models with each predictor removed.

The right-hand column of Figure 1 illustrates the combined formant shifts across all five vowels. Results of LMEM comparison indicated that noise had a significant effect on F1, increasing it by 1.6228 ± 0.0665 semitones ($p < 0.001$), and a significant but smaller effect on F2, increasing it by 0.2349 ± 0.0969 semitones ($p = 0.016$). There was also a significant effect of position upon F1, increasing it by 0.4523 ± 0.0665 semitones ($p < 0.001$), and upon F3, increasing it by 0.3190 ± 0.0819 semitones ($p < 0.001$). Significant effects were not found for vowel, subject sex or subject expertise. However, it is evident from Figure 1 that while the overall differences may not be significant, the increase of F1 in noise is relatively larger for the more close vowels [i, ɔ, u] than for open vowels [a, ɑ].

The results indicate that the noise in an MRI scanner has a greater effect upon F1 than position. The increase in F1 suggests that more open vowels are produced during MRI scans than in normal speech. This supports the findings of previous studies using broadband noise, including an articulatory study [19] which confirmed that the increase in F1 is due to a lowering of the jaw. The effect of noise upon F2 is smaller but still significant, and suggests that tongue position is affected by noise, independent of posture. As illustrated in Figure 2, the MRI noise has most energy in the region 1–2kHz, and future work will explore whether the spectrum of the MRI noise is linked to the formant shifts.

Speech produced in noise is also expected to show a decrease in spectral tilt, due to increased energy above 2kHz [8]. To determine the effect of MRI noise on speech, long term average spectra (LTAS) were calculated across all three read passages in each condition. Examples can be seen in Figure 2 for

Figure 2: Power-normalized LTAS (non-expert female subject), showing boosted frequencies where MRI noise energy is lower (e.g. ~3kHz).

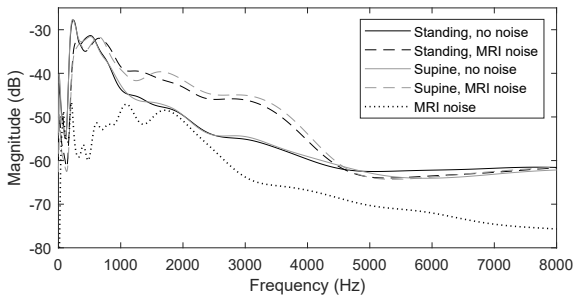
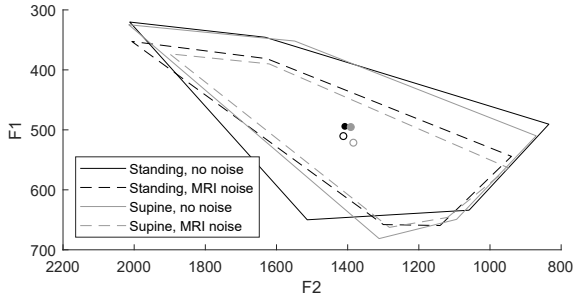


Figure 3: Vowel space differences (expert male subject), showing an approximately typical degree of reduction in vowel space size in noise.



a single subject, with dashed lines for the MRI noise conditions. The dotted line also illustrates the LTAS for the recorded MRI noise, and all five examples have been power normalized. It is clear from Figure 2 that for this speaker, the presence of MRI noise has a substantially greater effect on LTAS than position, suggesting that the speaker may have boosted speech frequencies (e.g. around 3kHz) where MRI amplitude is lower (a "bypass" strategy [8]). However, these results are highly subject-specific.

To quantify these differences, a LMEM was used to model spectral centroids on the range 0-8kHz, an approximate measure of spectral tilt. Noise condition, position, subject sex and expertise were used as fixed effects and subject as a random effect. Results indicated that spectral centroid increased by 26.91 ± 6.50Hz ($p < 0.001$) in noise, equivalent to an increase of 0.62% from the mean centroid in neutral conditions, with no significant effect of position.

It was suggested in [5] that the desire to be heard above noise may result in a larger vowel space, making vowels more distinctive. To explore this, vowel space dispersion—the mean Euclidean distance of vowel space vertices from the centroid—was calculated for each subject and condition, following [5]. Example vowel spaces can be seen in Figure 3 for a single subject. LMEMs were used with noise condition, position, subject sex and expertise as fixed effects, and subject as a random effect. Model comparison indicated that the vowel space dispersion decreased by 40.00 ± 4.90Hz^2 in MRI noise ($p < 0.001$) and a further 14.17 ± 4.90Hz^2 ($p = 0.005$) when the subject was supine. This equates to a reduction of 8.36% due to noise alone, and of 11.32% when the subject was also supine, compared to the mean vowel space dispersion in neutral conditions. The degree of this effect was subject-dependent, with some subjects showing a large difference across conditions, and some showing almost none. Based on feedback from some participants, it is suggested that disturbing the auditory feedback pathway made subjects less able to monitor the "correctness" of their speech and hence approach the corner vowel formant targets, thus reducing the vowel space dispersion, but this needs further investigation.

The results of this study indicate that the Lombard effect occurs in MRI-like noise conditions, and that the size of this effect considerably outweighs differences due to the supine posture required.

4. CONCLUSION

This study has shown that the Lombard effect occurs in MRI noise, affecting formant frequencies, vowel space dispersion and spectral tilt. Effects are both subject- and vowel-dependent. Data collection is ongoing to investigate these relationships.

In light of these results, the collection procedure for MRI vocal tract data and associated audio recordings should be carefully considered, and care must be taken when drawing conclusions from such data in isolation. If captured separately, audio data should be collected under conditions as close to the MRI setup as possible, including establishing consistency of communicative intent across conditions and using recorded scanner noise—which varies depending on the scanner and protocol used—at an appropriate amplitude. Future work will consider whether the effect varies between held vowels and running speech, whether there is an effect on laryngeal voice quality, and whether subject expertise affects the results, since "expert speakers" are commonly used for MRI data collection.

5. ACKNOWLEDGEMENTS

This research was funded by the Arts Humanities Research Council (AHRC) project Voice and Identity (AH/M003396/1).

6. REFERENCES

[1] Aalto, D., et al., 2014. Large scale data acquisition of simultaneous MRI and speech. *Applied Acoustics* 83, 64–75.

[2] Arnela, M., Guasch, O., Dabbaghchian, S., Engwall, O. 2016. Finite element generation of vowel sounds using dynamic complex three-dimensional vocal tracts. *Proc. 23rd Int. Congr. Sound Vib.* Athens, Greece.

[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [computer program] version 6.0.43. http://www.praat.org/. [Retrieved 8 September 2018].

[4] Cho, Z. H., et al., 1997. Analysis of acoustic noise in MRI. *Magnetic Resonance Imaging* 15(7), 815–822.

[5] Cooke, M., Lu, Y. 2010. Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *J. Acoust. Soc. Am.* 128(4), 2059–2069.

[6] Counter, S. A., Olofsson, A., Grahn, H. F., Borg, E. 1997. MRI acoustic noise: Sound pressure and frequency analysis. *J. Magnetic Resonance Imaging* 7(3), 606–611.

[7] Engwall, O. 2000. Are static MRI measurements representative of dynamic speech? Results from a comparative study using MRI, EPG, and EMA. *Proc. ICSLP 2000* Beijing, China. 17–20.

[8] Garnier, M., Henrich, N. 2014. Speaking in noise: how does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Comput. Speech Lang.* 28, 580–597.

[9] Garnier, M., Henrich, N., Dubois, D. 2010. Influence of sound immersion and communicative interaction on the Lombard effect. *J. Speech, Lang. Hearing Res.* 53, 588–608.

[10] Gully, A. J., Daffern, H., Murphy, D. T. 2018. Diphthong synthesis using the dynamic 3D digital waveguide mesh. *IEEE/ACM Trans. Audio Speech and Language Process.* 26(2), 243–255.

[11] Junqua, J.-C. 1993. The Lombard reflex and its role on human listeners and automatic speech recognizers. *J. Acoust. Soc. Am.* 93(1), 510–534.

[12] Lane, H., Tranel, B. 1971. The Lombard sign and the role of hearing in speech. *J. Speech Hearing Res.* 14, 677–709.

[13] Lombard, E. 1911. Le signe de l'elevation de la voix. *Annales des Maladies de L'Oreille et du Larynx* 37, 191–119.

[14] Lu, Y., Cooke, M. 2008. Speech production modifications produced by competing talkers, babble, and stationary noise. *J. Acoust. Soc. Am.* 124(5), 3261–3275.

[15] Peterson, G. E., Barney, H. L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175–184.

[16] Scott, A. D., Wylezinska, M., Birch, M. J., Miquel, M. E. 2014. Speech MRI: Morphology and function. *Physica Medica* 30(6), 604–618.

[17] Stowe, L. M., Golob, E. J. 2013. Evidence that the Lombard effect is frequency-specific in humans. *J. Acoust. Soc. Am.* 134(1), 640–647.

[18] Summers, W. V., et al., 1988. Effects of noise on speech production: acoustic and perceptual analyses. *J. Acoust. Soc. Am.* 84(3), 917–928.

[19] Šimko, J., Š. Beňuš, , Vainio, M. 2016. Hyperarticulation in Lombard speech: global coordination of the jaw, lips and the tongue. *J. Acoust. Soc. Am.* 139(1), 151–162.