DOI: 10.61782/fa.2023.0214

# THE EFFECT OF THE FREQUENCY ENERGETIC CONTENT OF NOISE ON THE LOMBARD EFFECT AND SPEECH INTELLIGIBILITY

**Pasquale Bottalico**[1*] **Silvia Murgia**[1]
[1] Department of Speech and Hearing Science, University of Illinois Urbana-Champaign, United States

## ABSTRACT

The Lombard Effect is an increase in vocal effort in response to rising noise levels and disturbance in the communication environment. The objective of this study is to evaluate the Lombard Effect during four intensity levels from 45dB(A) to 75dB(A) of low-frequency, medium-frequency, and high-frequency energy noise to measure the effect of disturbance and vocal discomfort on the speaker's intent to communicate, as well as the effect on speech intelligibility. Twelve conditions were randomly presented and recorded for each participant with the three types of noise and levels. At each condition, 20 participants were asked to read a passage. Immediately following each reading, participants were asked to rate the amount of noise disturbance and vocal comfort they had experienced. After, the speech intelligibility was evaluated by asking the participants to repeat the sentences of the QuickSin test emitted by a Head and Torso Simulator. The medium-frequency energy noise showed the highest Lombard Effect and the stronger decrease in intelligibility. In the conditions with noise with mid-frequency energetic content, the decrease in intelligibility was drastic with the increase in noise level. Low-frequency noises minimally impact speech intelligibility. High-frequency noises show little change in intelligibility with the increase in noise level.

**Keywords:** *Lombard Effect, speech intelligibility, frequency content, noise, communication disturbance, vocal comfort*

## 1. INTRODUCTION

Research has shown that the Lombard effect is a reflexive response to changes in the level of background noise, with speakers modifying their vocal effort to compensate for the degraded auditory input [1–5]. This results in Lombard speech, which is characterized by increased vocal amplitude and fundamental frequency, changes in first formant frequency, articulation, lengthening of vowel duration, and increased vowel intensity [6–12]. The mechanisms performed to modify vocal characteristics have been shown to make Lombard speech more intelligible than speech produced in the absence of noise [8].

The increase in vocal effort is characterized by the equivalent continuous A-weighted sound pressure level (SPL) of speech measured at a distance of 1 m in front of the mouth in anechoic conditions [13]. The vocal effort for average conversational speech is around 60 dBA at 1 meter from the mouth of the speaker, but this level increases as a function of the environmental noise level. When the conversational background noise is present at a level up to about 30-40 dB(A), speech is marginally affected with an increase of 0.24 dB(A)/dB(A) [5, 14, 15]. However, when the noise level exceeds about 43 dB(A), the average power of speech undergoes an increment of about 0.65 dB(A) for every unit increase in the noise level [16], until reaching the saturation at high noise levels due to physiological constraints ("ceiling effect") [17].

Although attention has been given to the Lombard effect and its effects in various conditions [3–5, 17, 18], no definitive knowledge is available regarding how the energetic content of the noise influences the Lombard effect. The human ear perceives frequencies between 20 and 20,000 Hz, but the most important frequencies for speech intelligibility are between 500 and 4,000 Hz [19]. In addition, the sensitivity of the human ear varies with frequen-

---

*Corresponding author: pb81@illinois.edu.

**Copyright**: ©2023 Bottalico et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

cies, with the highest sensitivity for frequencies critical for speech (1,000- 4,000 Hz) and less sensitivity above and below that range [20, 21].

Since speech self-monitoring depends on the perception filtered by the auditory system, and the auditory system sensitivity varies as a function of frequency, the energy content of the noises should have a different effect on the Lombard effect vocal response and the disturbance and discomfort that the noise at different frequencies produces. In particular, evidence has indicated that the vocal response in the Lombard effect is not generalizable to every competing sound in the environment. In fact, it has been shown to be sensitive to frequency content, particularly those critical for speech [18, 22]. Similarly, the different energy content of noise may also have a different effect on speech intelligibility. Noise with an acoustic spectrum similar to that of speech should result in more degraded speech perception. Consequently, the focus of this study was to determine whether there are different Lombard slopes (Voice level vs Noise level) when a broadband noise is characterized by a low (LF) (20-500 Hz), medium (MF) (500-4000 Hz) and high frequency (HF) (4000-20000 Hz) energetic content. We also focused on its relationship with perceived communication disturbance and vocal comfort, as well as speech intelligibility. The research questions were as follows:

1. Is there a difference in the Lombard Slope when the noise has an energetic content at low, medium, and high frequencies?
2. Is there a difference in the slope on the communication disturbance from noise when the noise has an energetic content at low, medium, and high frequencies?
3. Is there a difference in the slope on the comfort associated with noise level when the noise has an energetic content at low, medium, and high frequencies?
4. Is there a difference in the slope on the intelligibility associated with noise level when the noise has an energetic content at low, medium, and high frequencies?

Since the noise at medium frequency energetic content is the range where the hearing sensitivity is the highest and most of the information of the human speech is contained, we hypothesize that MF will have the most detrimental effect on the vocal effort, disturbance, and discomfort and will cause the highest reduction of speech intelligibility.

## 2. EXPERIMENTAL METHOD

### 2.1 Participants

The test was administrated to 20 participants between 18 and 32 years old with an average age of 22.4 (SD=3.9). The participants were equally distributed between males (10) and females (10). 17 participants were native speakers of American English while three were advanced speakers. Three participants underwent speech therapy at a young age and no one reported a history of hearing impairment. All participants signed informed consent for their participation in the experiment, which was approved by the Institutional Review Board of the University of Illinois Urbana-Champaign under Protocol No. 18179.

### 2.2 Room acoustics and Procedure

The experiment took place in a single-wall sound-proof booth where the participants were seated facing a human listener positioned 2.5 m away, to simulate a real communication setting. Two directional loudspeakers (KRK Systems studio monitor model Rokit5 G3) directed at the participant at 45° from the mouth axes, also placed at a 2.5 m distance, emitted broadband noises with energetic content at different frequencies: low frequencies (LF) (20 -500 Hz), medium frequencies (MF) (500-4000 Hz), and high frequencies (HF) (4000-20000 Hz). The spectra of the three types of noise are shown in Fig. 1

Twelve conditions were randomly presented and recorded for each participant with a combination of noise at 3 different frequency ranges and 4 levels with a step of 10 dB (45dB(A), 55 dB(A), 65 dB(A), and 75dB(A)). The noise levels for the twelve conditions were measured with the ears of a Head and Torso Simulator with Mouth Simulator (HATS, 45BC KEMAR, GRAS, Holte, Denmark), located in the participant seat in the booth and analyzed by means of NTI XL2 Audio and Acoustic Analyzer. At each condition, participants were asked to read a six-sentence excerpt of the Rainbow passage [23] which was displayed on a vertical screen in front of the participant. After that, the speech intelligibility was evaluated by asking the participants to listen and repeat the sentences of the QuickSin test [24] emitted by a HATS with a normal vocal effort of 60 dB(A) at 1 meter. [13] The order of the lists for the intelligibility test and the order of the noise conditions were randomized for each participant. The participants were asked to speak pretending they were talking to the person seated in front of them. The speech was acquired by a measurement microphone placed in front of the partic-

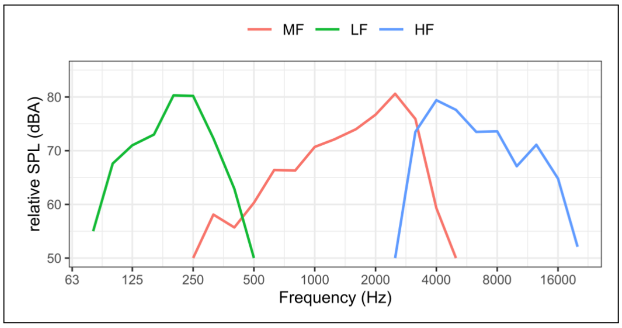

**Figure 1**. Spectra of the three types of noise with low, medium, and high energetic content.

ipant at mouth level at a distance of 15 cm. Immediately following each condition of noise, participants were asked to rate the amount of communication disturbance and vocal comfort, they had experienced. Participants marked their responses on a visual analog scale ranging from “Not at All” to “Extremely,” corresponding numerically to a range of 0 to 100.

The questions were the following:

1. Disturbance: How disturbed was your communication by the noise in this condition? (The extremes of the lines were “extremely disturbed” to the left and “not at all disturbed” to the right.)
2. Comfort: How comfortable was it to speak in this condition? (The extremes of the lines were “extremely” to the left and “not at all” to the right.)

The participant answered the questions by making a vertical tick on a continuous horizontal line of 100 mm length (a visual analog scale) right below each question.

### 2.3 Analyses

Analysis MATLAB (R2022a) was used for speech signal analysis. For each noise condition, the equivalent SPL was measured, and the mean value of the SPL was obtained per subject. For each subject, the average SPL among the conditions was computed and subtracted from each mean SPL value for that subject (termed $\Delta$SPL). This within-subject centering was performed in order to evaluate the variation in the subject’s vocal behavior in the different noise conditions from their typical vocal behavior (mean value of the SPL per subject). The levels were a combination of two sources: the voice and the noise. In order to evaluate the Voice to Noise Ratio (VNR) in the recordings, the distributions of the two sources were studied using the Expectation-Maximization algorithms for Gaussian mixtures [25, 26]. The algorithm allows for analyzing the mixture of distributions. In our case, the distribution of sound levels is a mixture of the voice and the noise levels. The algorithm estimates the mean values of the two distributions. The difference between the two mean levels represents an estimation of the VNR. The analysis was performed on a time history of the SPL, with a time step of 0.05 s, considering the subset of the dataset per noise condition.

Statistical analysis was conducted using the software R3.6.0 and the lme4 (version 1.1–10) package [27]. Linear mixed effect models were fit to the response variables $\Delta$SPL, self-reported disturbance, self-reported vocal discomfort, and intelligibility scores (IS) and as predictors, the noise level (Ln) (dBA), the type of noise (LF, MF, and HF), and their interaction. The listener ID was considered as a random factor. The self-rating of communication disturbance and discomfort was reported by each participant making a tick on visual analog scales. The score was obtained by measuring the distance between the left end of the line and the tick and converted in percentage from 0 (no disturbance or discomfort) to 100 (maximum disturbance or discomfort). The IS was measured as a percentage of words of the QuickSin test correctly identified for each acoustic condition. The models’ output includes the estimates of the fixed effects of the coefficients, the standard error associated with the estimate, the test statistic, t, and the p-value.

## 3. RESULTS

A total of 20 participants (n=10 male participants; n=10 female participants) have been tested by reading the rainbow passage and repeating the sentences pronounced by the HATS in twelve noise conditions (3 frequency ranges and 4 levels). As a first result, the VNR in the recordings was evaluated. The average VNR among the different noise conditions was 11.4 dB with a standard deviation of 3.6 dB. This result confirms that the effect of noise on the equivalent level was negligible. Model results of $\Delta$SPL and Ln are reported in Tab. 1. The relationship between SPL, grouped by noise at different frequency content, is shown in Fig. 2. The model showed a statistically significant relationship between $\Delta$SPL and Ln with $\Delta$SPL increasing when Ln increased for all the frequency ranges. Concerning the effect of noise type, there is a significant difference between the intercepts of the models for MF and LF, while the difference between the slopes is not significantly different. The intercepts of the two regression models for the MF and the HF are not significantly different while the difference between the slopes is approaching a statistically significant level with the slope for the MF noise 0.05 dB/dBA higher.

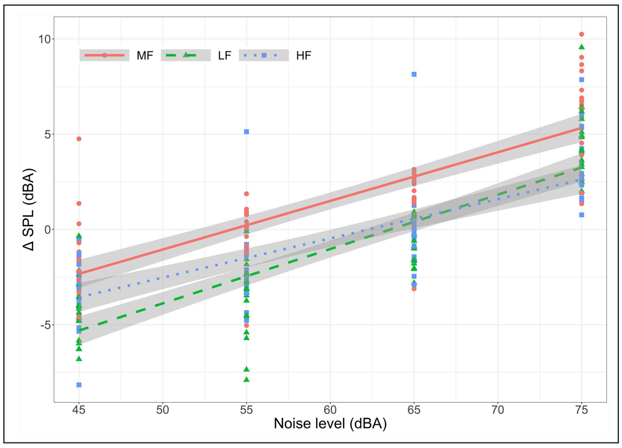

**Figure 2**. Regression lines from the model between Ln and ΔSPL. The results are grouped by frequency energetic contents, and the shaded regions indicate ± confidence interval.

Model results of the self-reported disturbance in the communication and Ln are reported in Tab. 1. The model showed a statistically significant relationship between disturbance and Ln, with the disturbance that increases when Ln increases for all the frequency ranges. Regarding the effect of noise type, there is no significant difference between the intercepts of the models for MF and LF, while the difference between the slopes is approaching a statistically significant level with the slope for the MF noise 0.48 dB/dB higher. The two regression models for the MF and the HF noise are significantly different in both intercepts and slopes. In particular, the model for HF has a higher intercept and a smaller slope compared to the model for MF.

Model results of the self-reported comfort in the communication and Ln are reported in Tab. 1. The model showed a statistically significant relationship between comfort and Ln with comfort that decrease when Ln increases for all the frequency ranges. Regarding the effect of noise type, there is a statistically significant difference between the intercepts and the slopes of both the models for MF and LF and MF and HF. In particular, the model for MF has a slope of 1.06 dB/dB higher compared to the model for LF, and 1.20 dB/dB higher compared to the model for MF.

**Table 1**. Piecewise Linear model output for four models with response variables ΔSPL, Disturbance, Time, and Budget as a function of Ln and Hearing Loss (with reference level equal to Normal Hearing).

| Predictor | Estimate | st.error | t-value | p-value |
|---|---|---|---|---|
| | | ΔSPL / dB | | |
| (Int.) | -13.82 | 1.22 | -11.30 | **0.001** |
| Ln | 0.26 | 0.02 | 12.74 | **0.001** |
| LF | -4.33 | 1.73 | -2.51 | **0.013** |
| HF | 0.99 | 1.73 | 0.57 | 0.567 |
| Ln:LF | 0.03 | 0.03 | 1.06 | 0.289 |
| Ln:HF | -0.05 | 0.03 | -1.74 | 0.083 |
| | | Disturbance / % | | |
| (Int.) | -37.83 | 11.78 | -3.21 | **0.002** |
| Ln | 1.80 | 0.19 | 9.60 | **0.001** |
| LF | -7.58 | 16.21 | -0.47 | 0.641 |
| HF | 34.60 | 16.21 | 2.13 | **0.034** |
| Ln:LF | -0.48 | 0.27 | -1.82 | 0.070 |
| Ln:HF | -0.78 | 0.27 | -2.95 | **0.003** |
| | | Comfort / % | | |
| (Int.) | 168.37 | 11.65 | 14.46 | **0.001** |
| Ln | -1.93 | 0.19 | -10.32 | **0.001** |
| LF | -38.18 | 16.10 | -2.37 | **0.001** |
| HF | -65.40 | 16.10 | -4.06 | **0.001** |
| Ln:LF | 1.06 | 0.26 | 4.00 | **0.001** |
| Ln:HF | 1.20 | 0.26 | 4.57 | **0.001** |
| | | Intelligibility Scores / % | | |
| (Int.) | 206.13 | 7.50 | 27.49 | **0.001** |
| Ln | -2.66 | 0.11 | -23.21 | **0.001** |
| LF | -85.03 | 9.90 | -8.58 | **0.001** |
| HF | -107.30 | 9.90 | -10.83 | **0.001** |
| Ln:LF | 2.17 | 0.16 | 13.34 | **0.001** |
| Ln:HF | 2.35 | 0.16 | 14.46 | **0.001** |

Model results of IS and Ln are reported in Tab. 1. The relationship between IS, grouped by noise at different frequency content, is shown in Fig. 3. The model showed

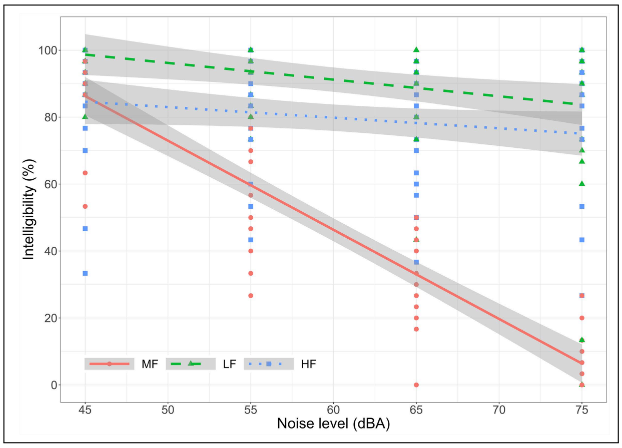

**Figure 3**. Regression lines from the model between Ln and intelligibility. The results are grouped by frequency energetic contents, and the shaded regions indicate ± confidence interval.

a statistically significant relationship between IS and Ln with IS that decreases when Ln increases for all the frequency ranges. Regarding the effect of noise type, there is a statistically significant difference between the intercepts and the slopes of the models for MF and LF and MF and HF. In particular, the model for MF has a slope of 2.17 %/dB higher compared to the model for LF, and 2.35 %/dB higher compared to the model for MF.

## 4. CONCLUSIONS

This study aimed to determine if broadband noise at different frequency ranges (low, medium, and high) produce different Lombard slope and have a different effect on perceived communication disturbance, vocal comfort, and speech intelligibility. The results showed that the medium frequencies generated the greatest increase trend in SPL (Lombard Effect) with the increase of the background noise level. A lower SPL, suggesting lower vocal effort, was instead produced with the noise in the low-frequency range, and when the noise was increasing, the rate of increase of voice was smaller compared to the other types of noise. The noise in the high-frequency range elicited similar vocal effort required as in middle frequencies at low noise levels but increased with a less steep slope compared to the conditions with medium frequencies noise. Regarding speech intelligibility, the increase of noise at mid-frequency energetic content produced the greatest and the steepest decrease in intelligibility. On the other hand, noise at low-frequency and high-frequency content had a minimal impact on speech intelligibility with small changes when the noise level was increased. The high-frequency background noise of 45 dBA seems to be enough to mask the high-frequency content of the speech spectrum at a normal vocal effort (i.e. 60 dBA at one meter in an anechoic condition).

## 5. ACKNOWLEDGMENTS

The authors would like to thank the participants for their valuable cooperation and interest.

## 6. REFERENCES

[1] C. L. Ludlow and D. B. Cikoja, “Is there a self-monitoring speech perception system?,” *Journal of communication disorders*, vol. 31, no. 6, pp. 505–510, 1998.

[2] E. Lombard, “Le signe de l’elevation de la voix,” *Ann. Mal. de L’Oreille et du Larynx*, pp. 101–119, 1911.

[3] H. Lane and B. Tranel, “The lombard sign and the role of hearing in speech,” *Journal of speech and hearing research*, vol. 14, no. 4, pp. 677–709, 1971.

[4] H. Lazarus, “Prediction of verbal communication is noise—a review: Part 1,” *Applied Acoustics*, vol. 19, no. 6, pp. 439–464, 1986.

[5] H. Lazarus, “New methods for describing and assessing direct speech communication under disturbing conditions,” *Environment International*, vol. 16, no. 4-6, pp. 373–392, 1990.

[6] K. D. Kryter, “Effects of ear protective devices on the intelligibility of speech in noise,” *The Journal of the Acoustical Society of America*, vol. 18, no. 2, pp. 413–417, 1946.

[7] J. Egan, “Psychoacoustics of the lombard voice response,” *Journal of Auditory Research*, vol. 12, no. 4, pp. 318–324, 1972.

[8] W. V. Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow, and M. A. Stokes, “Effects of noise on speech production: Acoustic and perceptual analyses,” *The Journal of the Acoustical Society of America*, vol. 84, no. 3, pp. 917–928, 1988.

[9] J.-C. Junqua, “The influence of acoustics on speech production: A noise-induced stress phenomenon known as the lombard reflex,” *Speech Communication*, vol. 20, no. 1-2, pp. 13–22, 1996.

[10] M. Södersten, S. Ternström, and M. Bohman, “Loud speech in realistic environmental noise: phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers,” *Journal of Voice*, vol. 19, no. 1, pp. 29–46, 2005.

[11] S. Ternström, M. Bohman, and M. Södersten, “Loud speech over noise: Some spectral attributes, with gender differences,” *The Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1648–1665, 2006.

[12] M. Garnier and N. Henrich, “Speaking in noise: How does the lombard effect improve acoustic contrasts between speech and ambient noise?,” *Computer Speech & Language*, vol. 28, no. 2, pp. 580–597, 2014.

[13] I. O. for Standardization, “Ergonomics–assessment of speech communication.” `https://www.iso.org/standard/33589.html`, 2003. ISO Standard No. 9921:2003.

[14] T. Korn, “Effect of psychological feedback on conversational noise reduction in rooms,” *The Journal of the Acoustical Society of America*, vol. 26, no. 5, pp. 793–794, 1954.

[15] M. Gardner, “Effect of noise, system gain, and assigned task on talking levels in loudspeaker communication,” *The Journal of the Acoustical Society of America*, vol. 40, no. 5, pp. 955–965, 1966.

[16] P. Bottalico, I. I. Passione, S. Graetzer, and E. Hunter, “Evaluation of the starting point of the lombard effect,” *Acta Acustica united with Acustica*, vol. 103, no. 1, pp. 169–172, 2017.

[17] H. Lane, B. Tranel, and C. Sisson, “Regulation of voice communication by sensory dynamics,” *The Journal of the Acoustical Society of America*, vol. 47, no. 2B, pp. 618–624, 1970.

[18] M. Garnier, N. Henrich, and D. Dubois, “Influence of sound immersion and communicative interaction on the lombard effect,” *Journal of Speech, Language, and Hearing Research*, vol. 53, pp. 588–608, 2010.

[19] N. French and J. Steinberg, “Factors governing the intelligibility of speech sounds,” *The Journal of the Acoustical Society of America*, vol. 19, no. 1, pp. 90–119, 1947.

[20] H. Fletcher and W. Munson, “Loudness, its definition, measurement and calculation,” *Bell System Technical Journal*, vol. 12, no. 4, pp. 377–430, 1933.

[21] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, and L. White, *Neurosciences*. De Boeck Supérieur, 2019.

[22] L. M. Stowe and E. J. Golob, “Evidence that the lombard effect is frequency-specific in humans,” *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 640–647, 2013.

[23] G. Fairbanks, *Voice and Articulation Drillbook*. Harper & Row, 2 ed., 1960.

[24] M. C. Killion, P. Niquette, G. Gudmundsen, L. Revit, and S. Banerjee, “Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners,” *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2395–2405, 2004.

[25] D. D’Orazio, D. De Salvio, L. Anderlucci, and M. Garai, “Measuring the speech level and the student activity in lecture halls: Visual-vs blind-segmentation methods,” *Applied Acoustics*, vol. 169, p. 107448, 2020.

[26] D. De Salvio, D. D’Orazio, and M. Garai, “Unsupervised analysis of background noise sources in active offices,” *The Journal of the Acoustical Society of America*, vol. 149, no. 6, pp. 4049–4060, 2021.

[27] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting linear mixed-effects models using lme4,” *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.