Modelling the Lombard Effect Using Multivariate Regression

Aim: To model the relationship between environmental noise level and vocal response using a supervised regression approach, demonstrating the application of basic multivariate machine learning techniques to experimental acoustic data.

This analysis extends a prior exploratory study of the Lombard effect by introducing a data-driven regression model implemented in PyTorch. The objective is not prediction or deployment, but to demonstrate feature construction, model training, and interpretation under realistic data constraints.

Model Behaviour (Predicted vs Observed plot):

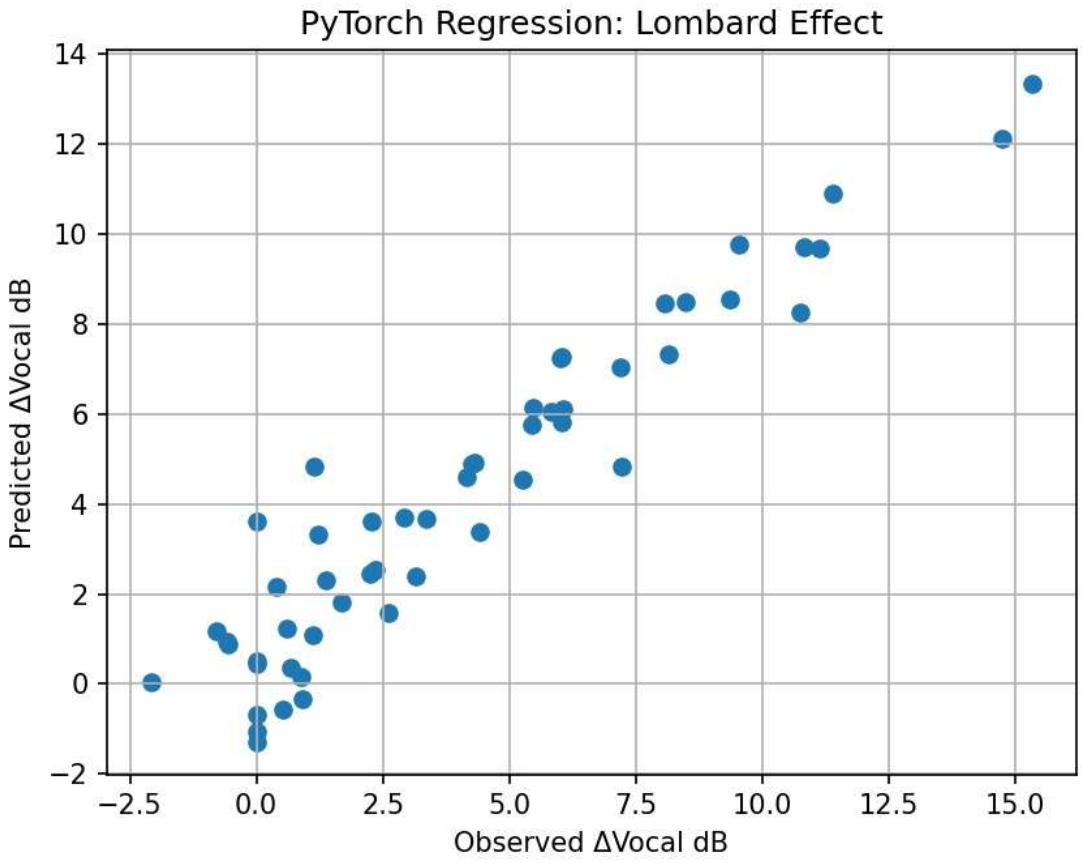

Figure 1: Comparison between observed and model-predicted changes in vocal intensity relative to baseline. The clustering around the diagonal indicates that the regression model captures the dominant trend linking environmental noise level to vocal response, despite inter subject variability

- The model reproduces the monotonic increase in vocal intensity with increasing noise level
- Deviations from the diagonal reflect participant-dependent response differences
- The spread increases at higher vocal effort, suggesting non-uniform saturation behaviour
- The model is trained on normalized relative dB values rather than absolute SPL

Learning Dynamics (Loss curve):

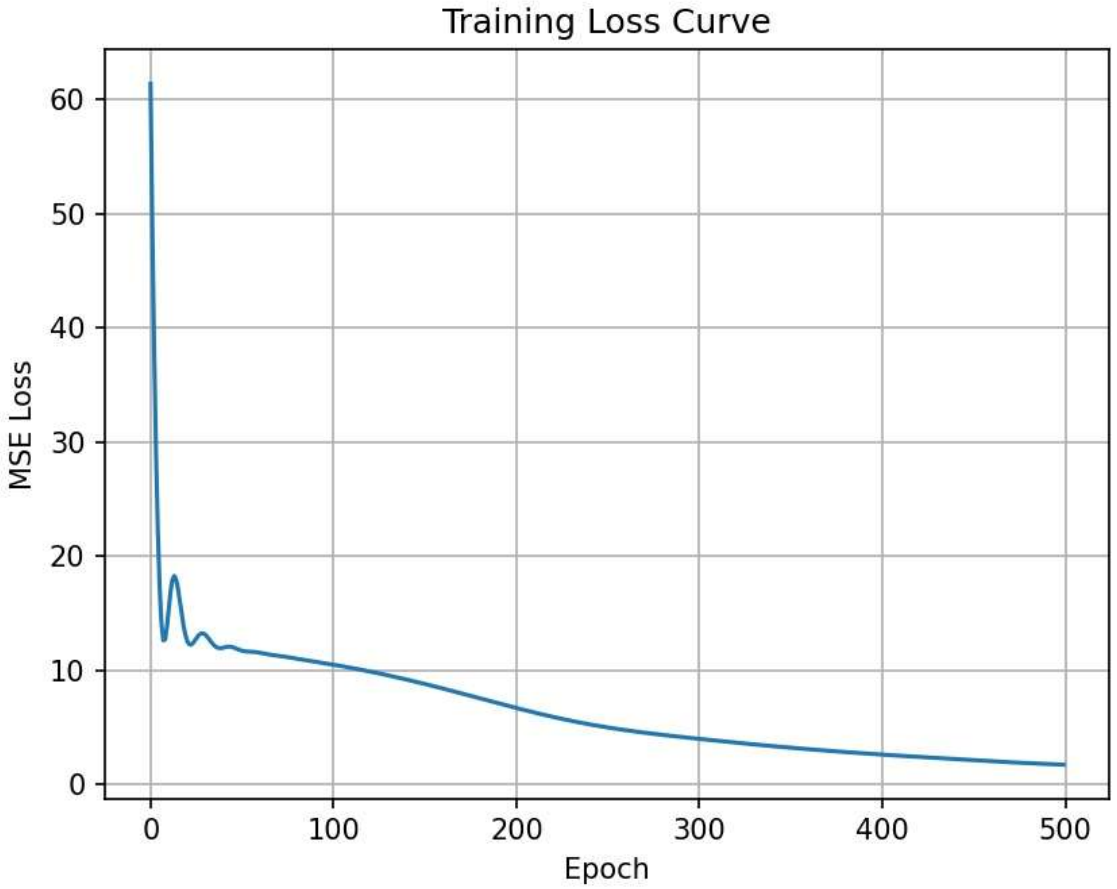

Figure 2: Training loss (mean squared error) as a function of epoch during model optimization.

- Rapid initial loss reduction indicates capture of the dominant Lombard trend
- Early oscillations are consistent with small dataset size and inter-subject variability
- Gradual convergence suggests stable optimization rather than overfitting or divergence.

Given the limited dataset and absence of a held-out test set, the model is not intended to demonstrate generalization performance. Instead, it serves as a proof-of-concept for applying supervised regression techniques to experimental acoustic data.