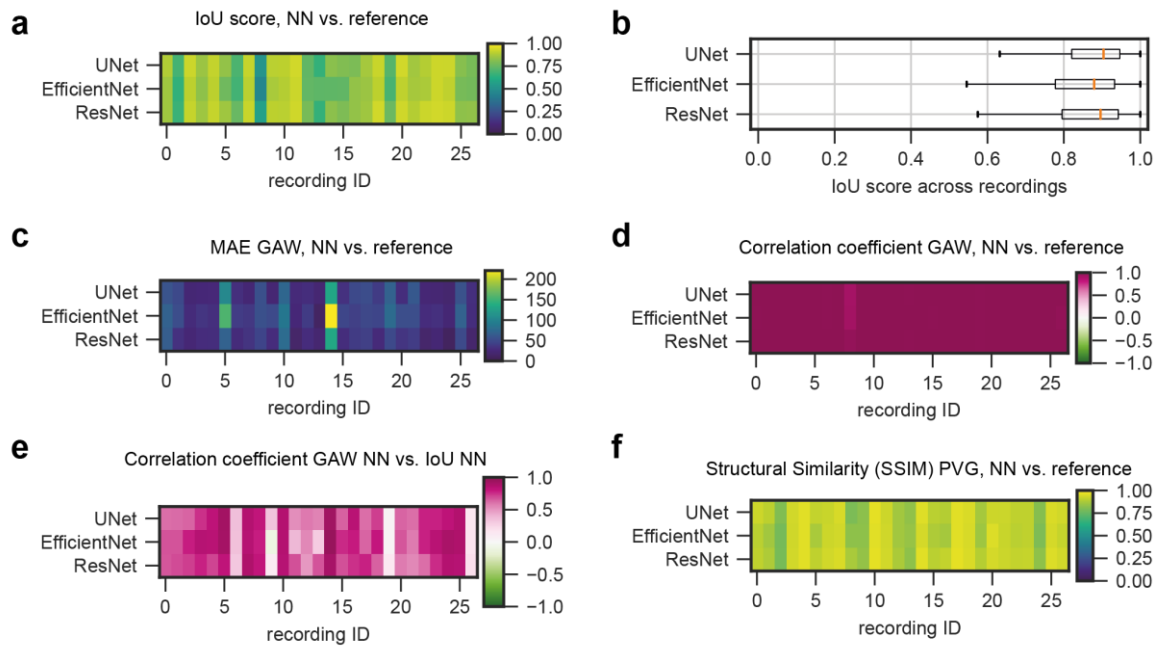**Supplemental Figure S1. Neural network training metrics. a)** Intersection over Union (IoU) plots across training epochs for custom U-Net (orange), EfficientNetB0 (green) and ResNet-50 (blue) for training (solid line) and validation (dashed line) dataset. **b)** Dice loss plots across training epochs for custom U-Net (orange), EfficientNetB0 (green) and ResNet-50 (blue) for training (solid line) and validation (dashed line) dataset. **c)** K-fold cross-validation of deep neural networks using different training/validation splits (k = 6). Blue crosses indicate the neural networks that are used in GAT.

Supplemental Figure S2. Neural network evaluation. a) Average IoU score across neural networks compared to reference (semi-automatic segmentation). **b)** Distribution of IoU-Scores, computed on previously unseen data. **c)** Mean absolute error (MAE) of the neural network glottal area waveforms (GAWs) compared to the reference GAW. **d)** Correlation coefficient between the neural network GAWs and the reference GAW. **e)** Correlation coefficient between neural network GAW and IoU score. **f)** Structural similarity (SSIM) between the neural network phonovibrogram (PVG) and the reference PVG. High values indicate high similarity.
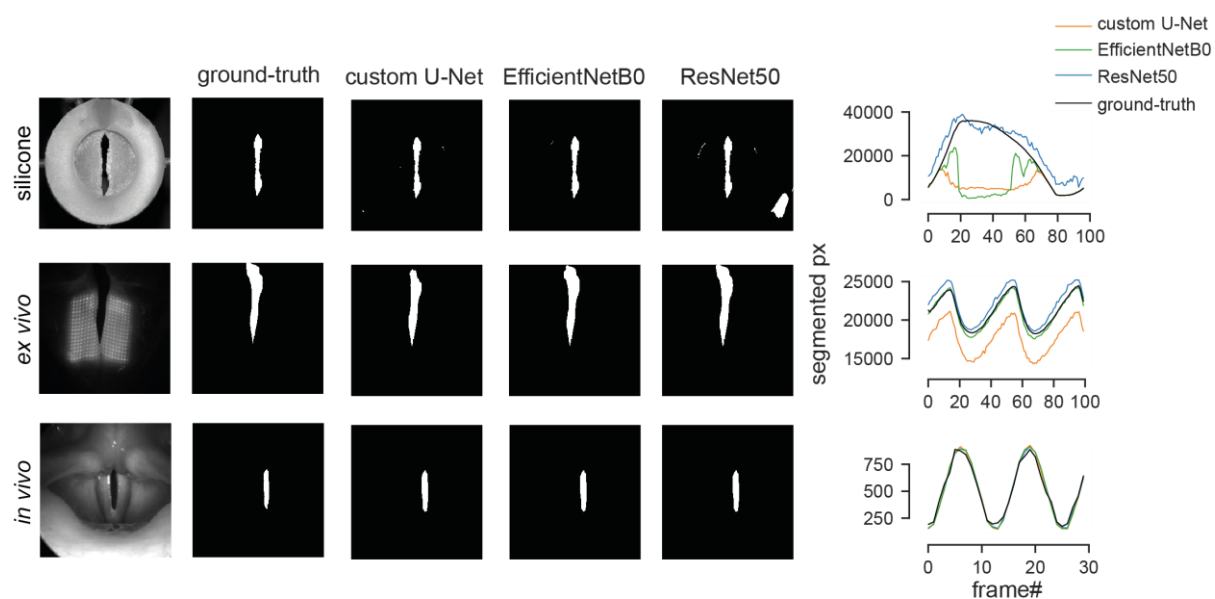
Supplemental Figure S3. Example segmentations using different backbones from various source data. From top to bottom: silicone vocal folds, *ex vivo* recordings, *in vivo* recordings.
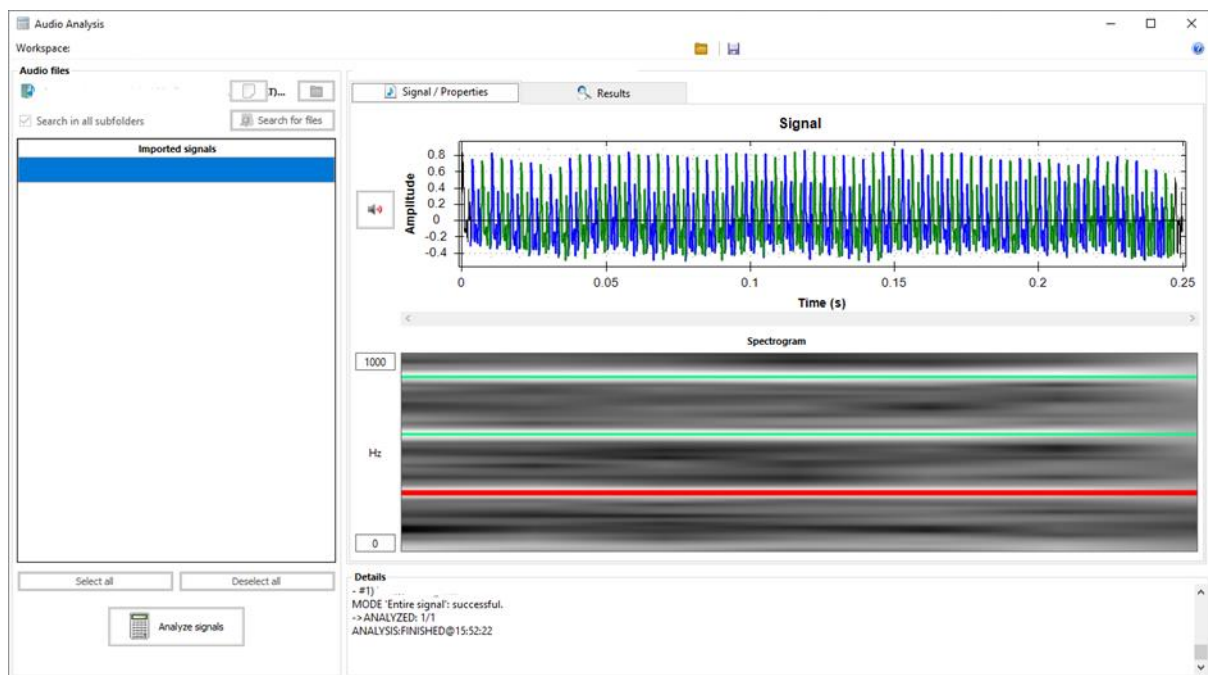
Supplemental Figure S4: Audio analysis platform. We provide next to the raw signal a spectrogram and a comprehensive parameter analysis.

Supplementary Movie 1: Endoscopic footage with semi-automatic and fully-automatic segmentation methods, multiple examples.

Supplementary Data: GAT Manual