Vocal Rise Time (VRT) measurement for Vowel Onset

Optimal algorithm selection for VRT signal envelope generation

John Holik[1], Brian Stasak[1], Duy Duong Nguyen[1], Tünde Szalay[1], Tomás Arias-Vergara[2], Michael Döllinger[3] and Catherine Madill[1]

[1]Sydney Voice Lab (Dr Liang Voice Program), University of Sydney [2]Pattern Recongnition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
[3]Division of Phoniatrics and Paediatric Audiology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

Introduction

Vocal Rise Time (VRT) is an acoustic-amplitude measurement of initial production of a vowel. It is utilised by clinicians and researchers alike to categorise glottal attack as soft, hard and breathy[a]. VRT can vary depending on the speaker and recording conditions[b].

This work used the VOAT[c] VRT measurement software to explore multiple envelope generation algorithms with a range of fitting parameters. These generate audio signal envelopes (example in Figure 1) to then find the difference between signal vowel onset start and end points (filled circles) relative to the vowels eventual maximum amplitude (star).

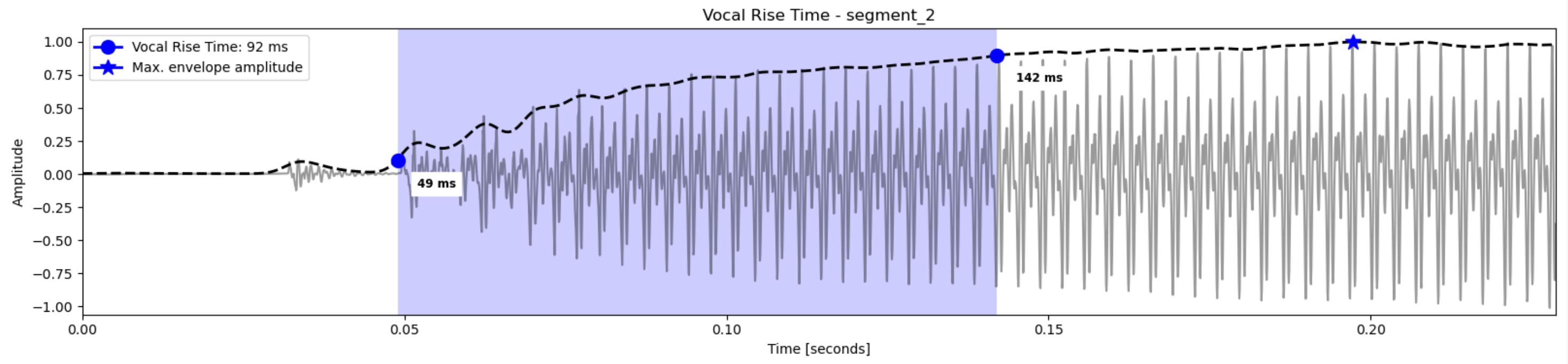


Figure 1: VRT measurement taken using VOAT software. Onset and saturation points (circles) and maximum amplitude (star).

[a]Maryn, Y & Poncelet, S - How Reliable Is the Auditory-Perceptual Evaluation of Phonation Onset Hardness? - Journal of Voice, 35(6) (2021)
[b]Chacon, A.M. - Vowel onset measures and their reliability, sensitivity and specificity: A systematic literature review - PLOS ONE, 19(05) (2024)
[c]Arias-Vergara, T - VOAT: Voice Onset Analysis Tool - SoftwareX (2024)

Which envelope to use?

For signal envelope generation, the VOAT has three algorithms with six algorithm specific fitting parameters to choose from. The selection of an optimal combination of the algorithm and fitting parameter is ideally done by visual appraisal of the envelope, but this is impractical with large numbers of files. It is necessary to ensure VRT measurement is not adversely affected by such things as noise, vocal dynamics and voice type, etc.

- Which combination of algorithm and fitting parameter is reliably optimal for measuring the VRT?

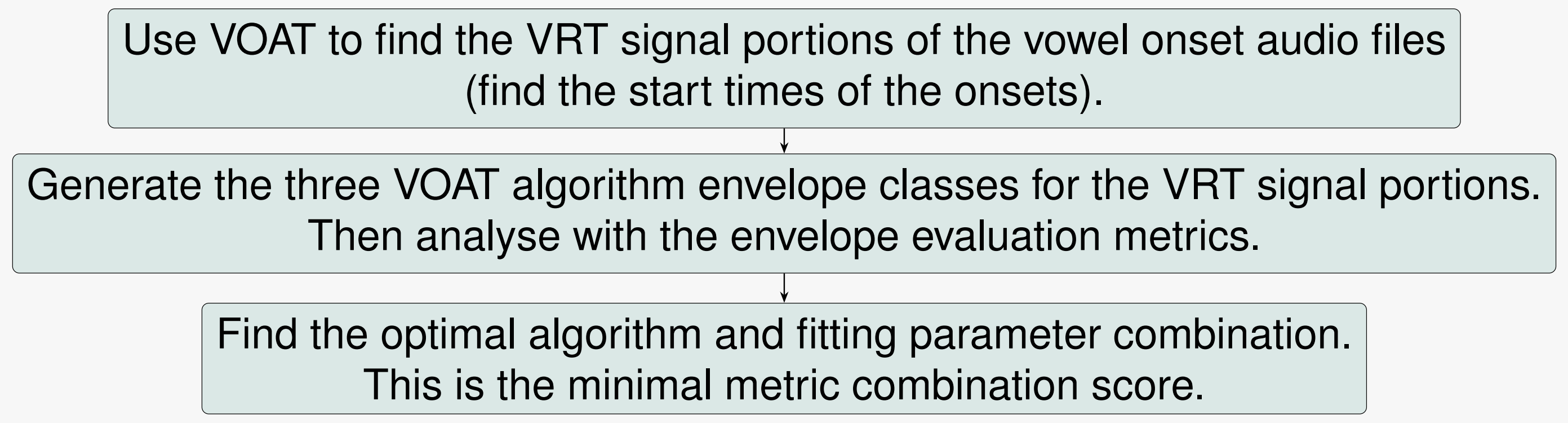
Dataset and VRT Extraction

Dataset: this dataset comprised 114 female adults (18-40 years, mean 23 ± 4) who self-reported as being healthy and having no symptoms of voice disorder. They produced three repetitions of a sustained vowel /a/ in Australian English.

University of Sydney, Human Research Ethics Committee (Approval No. [2016/1001]).

VRT Extraction: The primary data were run in VOAT and the VRT portions identified. The envelopes of these portions were then generated with the VOAT algorithms. These envelopes then comprised the datasets measured using the metrics outlined in Table 1.

VRT Experimental Process Flowchart



Use VOAT to find the VRT signal portions of the vowel onset audio files (find the start times of the onsets).

Generate the three VOAT algorithm envelope classes for the VRT signal portions. Then analyse with the envelope evaluation metrics.

Find the optimal algorithm and fitting parameter combination. This is the minimal metric combination score.

Envelope Generating Algorithms

These VRT signal portion envelopes are presented below in Figures 2-4. These start 20 ms before the vowel onset start point.

Hilbert

The Hilbert method starts with taking the Hilbert transform of the source signal. It is then convolved with a Gaussian curve of width 'Window Length'. These fitting parameter Window Lengths are 0, 40, 80, 120, 160, 200 ms. The longer the window length the wider the convolving Gaussian curve and hence smoother envelope.

- Signal amplitude dominates this envelope as the Gaussian curve of varying windows is 'dragged' over the instantaneous signal amplitude.

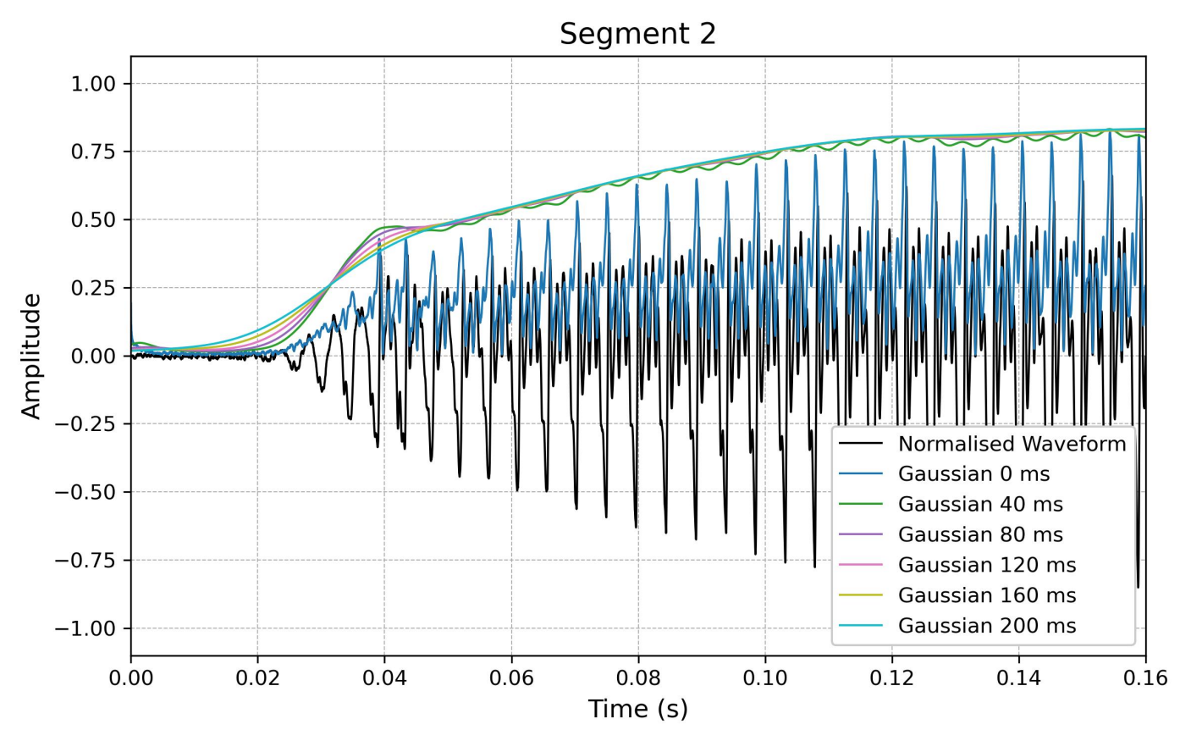


Figure 2: Hilbert envelopes for the second /a/ of participant A.

Peak Amplitudes

The Peak Amplitude method performs peak detection over the band-pass filtered audio file, then generates the envelope by joining those points. The fitting parameter here is the 'Smoothing Factor', where higher numbers increase the distance between peaks, hence a smoother envelope. For these audio recordings, Smoothing Factors 1 & 2 are almost identical.

- Distance between peak amplitudes dominate this envelope.

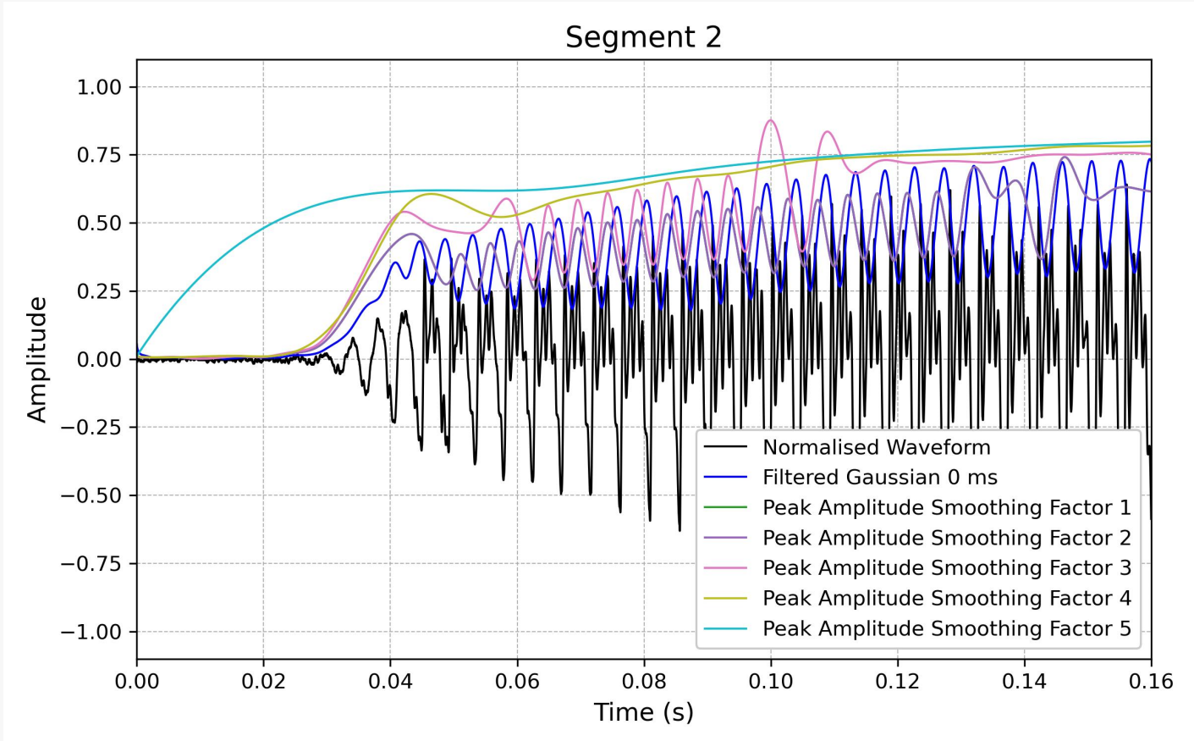


Figure 3: Peak Amplitudes envelopes for the second /a/ of participant A.

Root Mean Square (RMS)

The Root Mean Square method takes the square root of the squared mean of the source values within the fitting parameter 'Window Length'. The Window Lengths are 0, 20, 40, 60, 80, 100 ms. The wider the Window Length, the smoother the envelope.

- Signal energy dominates this envelope as the window length defines the region over which the signal is averaged.

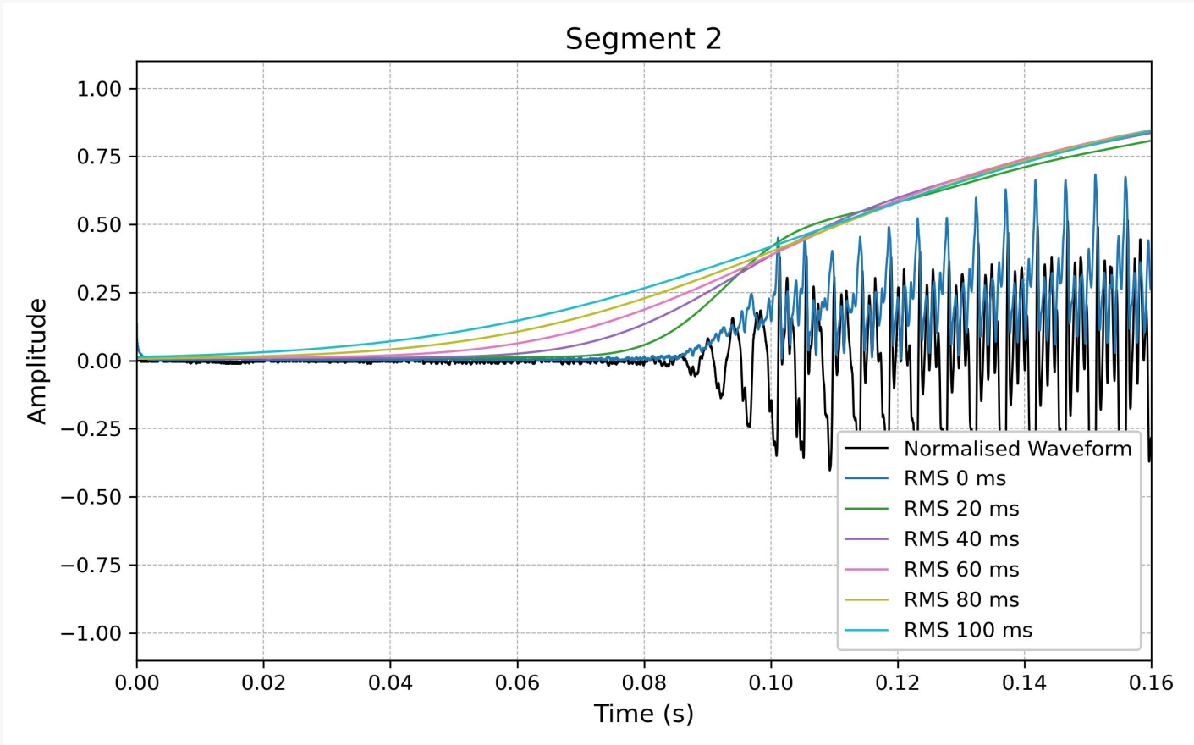


Figure 4: Root Mean Square envelopes for the second /a/ of participant A.

How to choose the optimal envelope?

In Figures 2 & 4, the envelopes for Hilbert and RMS are overly smooth. This demands the envelope evaluation metrics be chosen to favour less flat envelopes. Contrary to this, the Peak Amplitude envelopes are overly 'curvy' (discard Smoothing Factor 5 envelopes as these consistently start too early & are overly smooth), so require the envelope evaluation metrics to penalise for this.

This motivates the selection of tailored envelope comparison and curve description metrics combinations. These are the 'Optimal Envelope Score' values seen in Figures 5 - 7, with the metric combination equation displayed in the bottom of each plot.

Optimal Envelope Score and Envelope Evaluation Metrics

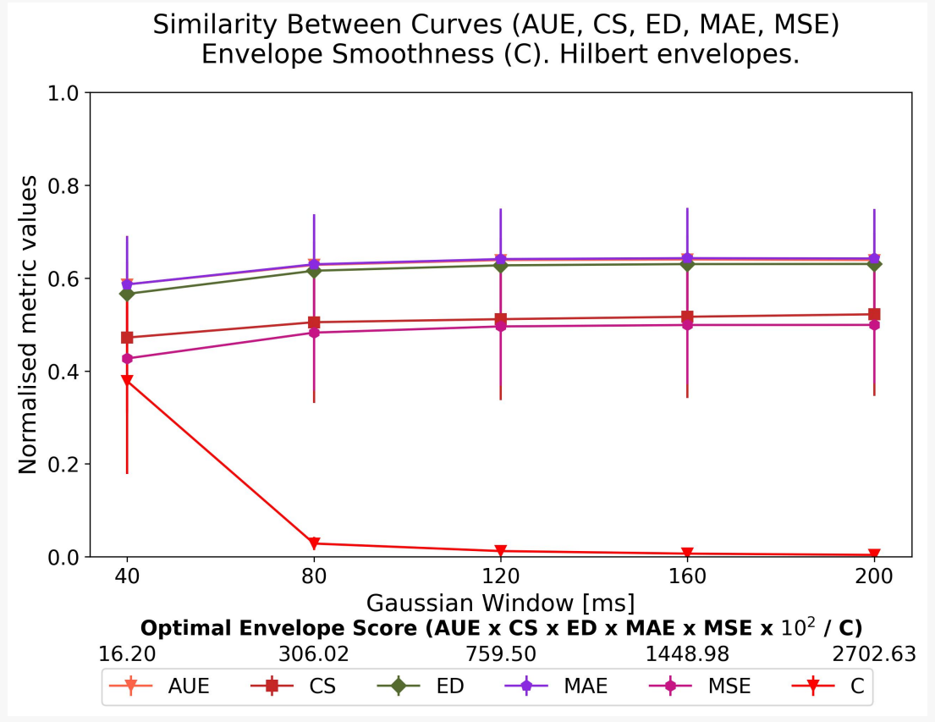


Figure 5: Hilbert metrics. Less smooth envelopes favoured for this combination → division by smoothness metrics.

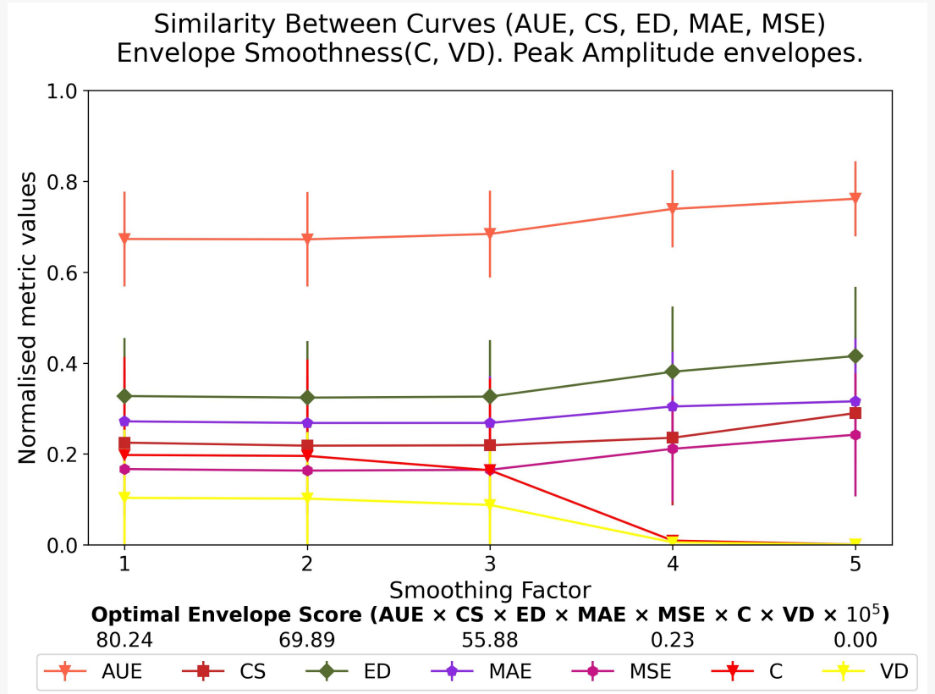


Figure 6: Peak Amplitude metrics. Smooth envelopes favoured → multiplication of smoothness metrics + scaling factor.

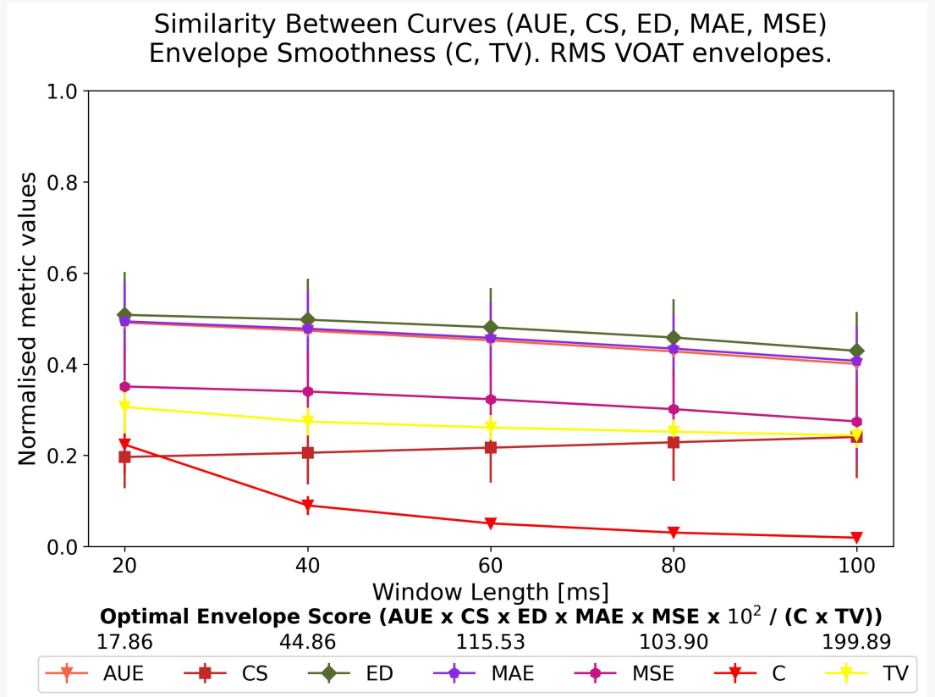


Figure 7: RMS metrics.. More variable envelopes favoured for this combination → division by smoothness metrics.

Optimal Envelope Score

Figures 5-7 plot the metric values for each fitting parameter, averaged over all 114 participants. The Optimal Envelope Score (OES) for each fitting parameter is an algorithm-specific combination of the Table 1 metrics: the lower the value, the better the envelope fit to the audio signal.

Envelope Evaluation Metrics

Similarity between curves: These measure differences between the source signal and the envelope curves. Lower values → more similar curves.

Envelope Smoothness: These measure a morphological aspect of the envelope. Lower values → smoother curves.

Similarity Between Curves Metrics	
Definition	**Equation**
Area Under the Envelope (AUE): envelope area minus source signal area.	$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$
Cosine Similarity (CS): measures directional alignment of the curves. Can capture differences in curve gradients. (similar curves approach 1, so as-used metric: (1 - CS)).	$\frac{X.Y}{\lVert X\rVert\,\lVert Y\rVert}$
Euclidean Distance (ED): measures element-wise distance between curves.	$\sqrt{\sum(x_i - y_i)^2}$
Mean Absolute Error (MAE): measures mean errors between paired observations. Less sensitive to outliers.	$\frac{1}{n}\sum \lvert x_i - y_i\rvert$
Mean Squared Error (MSE): measures the mean squared errors between paired observations. High sensitivity to outliers.	$\frac{1}{n}\sum(x_i - y_i)^2$
Envelope Smoothness Metrics	
Curviness (C): finds the absolute value of the second derivative of the curve. How much the slope itself is changing.	$\frac{1}{n-2}\sum_{i=1}^{n-2}\left\lvert \frac{d^2y}{dx^2}_i \right\rvert$
Total Variance (TV): sums the absolute values of the first derivative. Captures the overall variation or 'roughness'.	$\frac{1}{n-1}\sum_{i=1}^{n-1}\left(\frac{dy}{dx}_i - \overline{\frac{dy}{dx}}\right)^2$
Variance of first Derivative (VD): measures how much the slope fluctuates.	$\mathrm{Var}\left(\frac{dy}{dx}\right)$

Table 1: Envelope evaluation metric definitions and equations. Envelope evaluation metric details: X (x_i) and Y (y_i) represent the envelope and the source curves (individual elements) respectively.

Conclusions

- An Optimal Envelope Score (OES) of the signal envelopes is proposed.
- For each algorithm, the optimal fitting parameter for this dataset are: Hilbert 40 ms, Peak Amplitude SF 4 and RMS 20 ms. These are the best settings to use as they produced each algorithms lowest OES (Peak Amplitude SF 5 excluded as an outlier).
- When approaching the VRT measure for a new dataset, the nature of the signal must be taken into account. The Hilbert is likely the best for most applications. Some voice onset types may benefit from different envelope algorithms to differentiate pathological from normative voice types.

Future works are to explore incorporating these envelope evaluation metrics into an iterative optimisation routine. This would allow for optimal algorithm and fitting parameters to be automatically found for each new dataset.

Sydney Voice Lab 16th Advances in Quantitative Laryngology, Voice and Speech Research — Groningen, June 24 – 27, 2025 john.holik@sydney.edu.au