Frication Peak Landmarks

J. MacAuslan

Landmarks (LMs) are acoustically identifiable points in an utterance. They come in the form of abrupt transitions (*abrupt* LMs) and peaks (*peak* LMs) of some contour or contours.

Until now the only peak type has been Vowel, computed by *vowel_lms*. For vowels the peak is that of maximum harmonic power and often corresponds to the maximum opening of the mouth.

Frication-type peak landmarks are computed using *fricative_lms*. Frication or aspiration is necessarily characterized by weak energy at low frequencies and strong at high frequencies. A few kHz is a typical frequency for frication in human speech.

Strong frication is generated by well-developed air turbulence. The fractal dimension of *all* well-developed turbulence is 5/3 when measuring scalar signals such as pressure. This remarkable fact is the consequence of a universal property of turbulence, the Kolmogorov spectrum.

A fractal dimension can be thought of as the degree of roughness or smoothness in the graph or plot of the waveform over time. The smoothest possible signal will have a fractal dimension of one, a smooth curve. The roughest possible signal will fill the entire two dimensional plot area and have a fractal dimension of two.

Perhaps surprisingly, computing fractal dimension is fast and simple. It depends solely on the difference between the maximum and minimum values of the signal over 2, 4, 8, and 16 samples (for instance).

We compute a fractal dimension contour instant by instant over intervals of approximately one msec. Therefore, the fractal dimension measures the dimensionality of the signal around 2 kHz (8 samples at 16 kHz), ensuring that multiple measurements are available for comparison at a timescale of approximately ½ msec.

The primary indicator of the peak of frication is a fractal dimension as close to 5/3 as possible. However, if a signal has a fractal dimension of 5/3 but has high power in low frequencies, we do not consider this a fricative. It may be air turbulence but it cannot be assumed to be a component of speech. Fricative turbulence is produced by a narrow vocal-tract constriction, which automatically imposes a reduction in energy at low frequencies.

Brief and voiced frication both often have fractal dimensions somewhat less than 5/3, typically approximately 1.5. In contrast, vowels have fractal dimensions between 1 and 1.2 (1 being the lowest possible value for fractal dimensionality). Artificial signals including those of clipped audio can have fractal dimensions close to 2, the upper limit.

The figures show examples of the fractal-dimension and high-vs.-low frequency energy contours defining frication as well as the specific points – the landmarks -- that are identified as the local peaks of those contours. In Figure 1, notice that the stop release is so strong (perhaps hyper-articulated) that the resulting burst shows fully developed turbulence and an F landmark. As is true throughout the SpeechMark® suite, this tool analyzes the speech *as produced*, not necessarily as expected or even as intended.

Figure 1 illustrates one other point: These two contours may have meaningless values during near-silent intervals. Therefore, all processing suppresses detection during such intervals, to avoid finding F LMs before and after the production. (This is accomplished with a simple total-energy contour, not shown in the figures.)

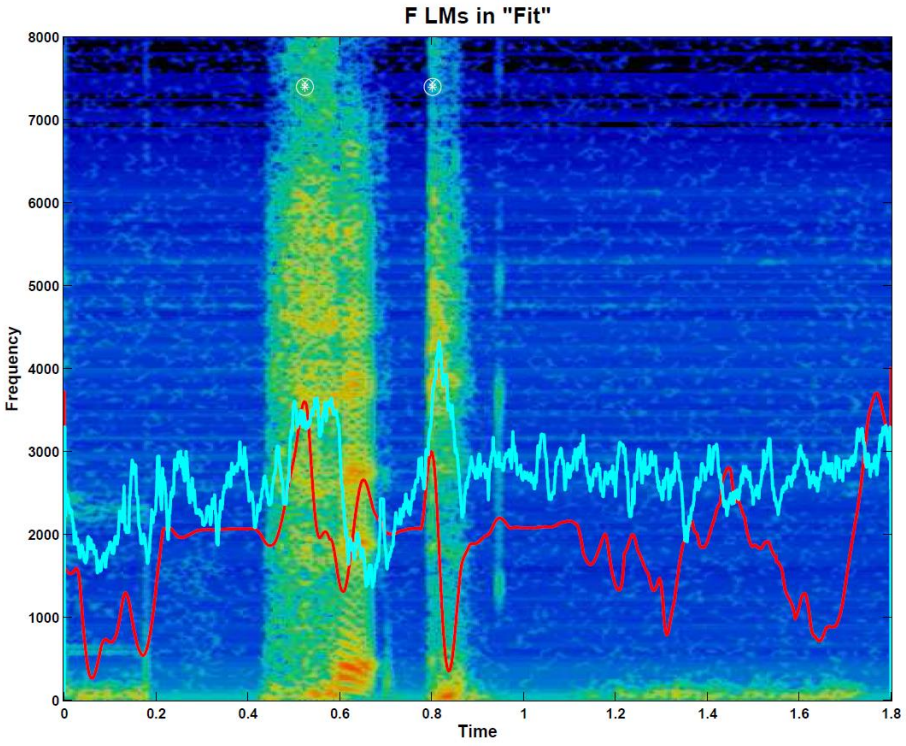

Figure 1. Spectrogram, Scaled Contours and Frication (F) Landmarks in "Fit". Note the vowel produced during 0.60-0.65s and the prominent stop release at 0.80-0.85s. (*light blue*) Contour of fractal dimension. (Scaling: 0 => dimension 1, 4000 => dimension 5/3.) (*red*) Contour of energy contrast, high-frequency – low-frequency. (Scaling: 0 => LF dominant; 4000 => HF dominant.) *Not shown*: total-energy contour to suppress detection in near-silent intervals. (*white marks* shown at 7500 Hz) F landmarks at 0.52s, 0.80s.

In Figure 2, we see the importance of using both the dimension and contrast contours. For example, strong contrast peaks occur at 2.30s (*air*) and 2.75s (*act*), due to sonorants with high energy at high frequencies; however, the fractal dimension is appropriately very low there, so no landmarks are detected. Conversely, at 2.9s (*act*) and 3.1s (*like*), strong though brief bursts produce fractal-dimension peaks; but with low contrast between high- and low-frequency energy, these do not produce landmarks.

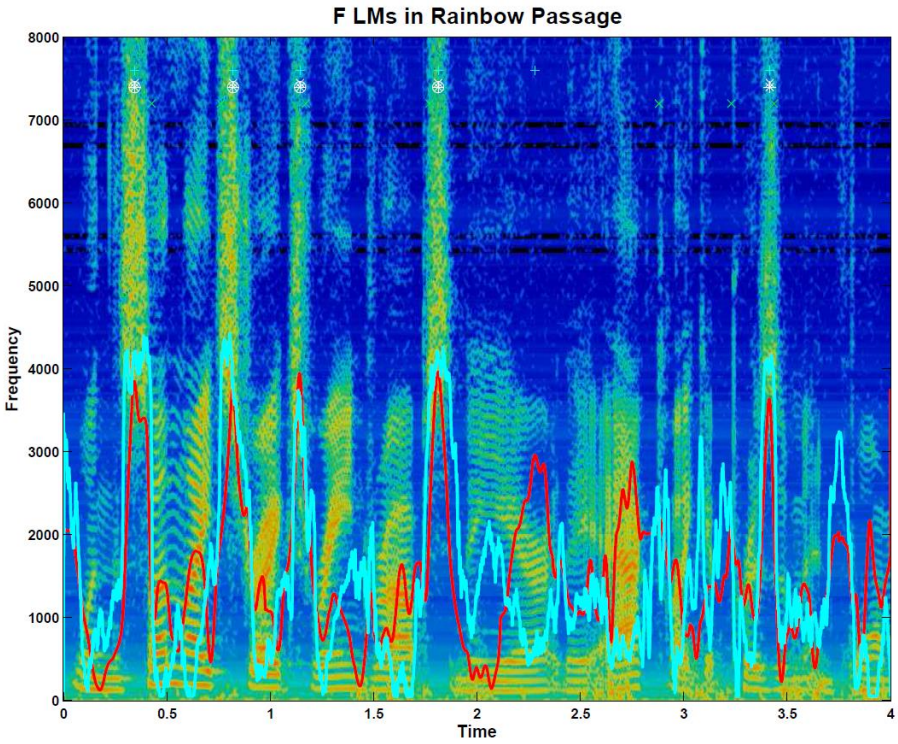

Figure 2. Spectrogram, Scaled Contours and Frication (F) Landmarks in Sentence. The sentence is "When the sunlight strikes raindrops in the air, they act like a prism and form." Contours and F landmarks as in Figure 1. Note that both voiced stridents (*priSm*, 3.4s) and unvoiced ones (*Sunlight*, 0.4s; *StrikeS*, 0.8 and 1.1s; *raindropS*, 1.7s) have sufficient airflow and duration for well-developed turbulence, whereas other points (*the*, 0.3s; *form*, 3.8s) and bursts (*act*, 2.9s; *like*, 3.1s) may not.