# NCMF Methodology for Human Assisted Speaker Recognition

*Approximate execution time*: 4 hours per trial with negligible amount of machine processing time.

1. Convert two-channel .sph files to mono 8 kHz 16 bit .wav files.
    a. Note target speaker's channel
2. Build speaker profiles based on aural evaluation of voice characteristics for both model and test segments. Listen for:
    a. Voice quality
    b. Dialect/accent articulation
    c. Patterns in pronunciation
    d. Pathological patterns
    e. Perceived pitch
    f. Perceived rate
    g. Perceived amplitude
    h. Breath patterns
    i. Syllabic patterns
3. Edit auxiliary speaker and extraneous noises from files. Save as edited 8 kHz and 11 kHz files.
4. Extract acoustic parameters from edited files using:
        i. Multispeech Model 3700 Version 3.1.7 www.kayelemetrics.com
        ii. Matlab R2010a www.mathworks.com
        iii. Wavesurfer Version 1.8.5 http://www.speech.kth.se/wavesurfer/
        iv. Catalina Version 3.0h http://www.forensicav.ro/download.htm
    b. Parameters include:
        i. Average and standard deviation of F0 in Hz
        ii. Average spectral power between 50 Hz-300 Hz, average spectral power from 1 kHz-3.5 kHz, and the ratio of these values.
        iii. Standard deviation of energy in dB
        iv. Average value and standard deviation in Hz of F1, F2, and F3 separately for entire voice sample
        v. Average value in Hz for Formants 1, 2, and 3 for vowels a, e, i, o, and u separately
5. Assess parameter-to-parameter deviation percentages between 'model' and 'test' segments. Flag those outside a given range established through research at NCMF.[1]
6. Considering those parameters not flagged as identity matches between 'model' and 'test' segments, calculate likelihood ratio as the probability of the evidence given the two hypotheses:

$$LR = \frac{p(E|H_t)}{p(E|H_f)} \quad -or- \quad LR = \frac{matched\ parameters}{unmatched\ parameters}$$

[1] *Forensic Voice Identification Utilizing Digitally Extracted Speech Characteristics*, J.M. Smith and R.W. Sanders. Presented at the 125th Convention of the Audio Engineering Society, San Francisco, 2008.

7. Finally, a panel of 4 UC-Denver graduate students enrolled in the course **MSRA 6530: Graduate Audio Forensics** was utilized to provide their decision for each trial. The members of the panel would aurally review the model and test segment audio samples along with the computationally derived *LR* and provide their 'true' or 'false' conclusion and a subjective "percent of confidence" of that decision. These decisions were averaged with the *LR* to derive a combined likelihood ratio (*CLR*). The $\log_{10}$ of the *CLR* was provided to NIST for each trial as well as the decision of 't' or 'f'.