**Part 1: Feature Extraction**

1. **What audio features did you choose to extract and why?**

* Mel Cepstral Coefficients(MFCC’s) and LPC coefficients are selected as features of audio speech for child and parent, Because sounds generated by a human are filtered by the shape of the vocal tract including tongue, teeth etc. This shape determines what sound comes out.
* The shape of the vocal tract manifests itself in the envelope of the short time power spectrum, and the job of MFCCs is to accurately represent this envelope.
* LPC help to add pitch information to the features.

1. **Are there any features you intentionally didn't use?**

* I ignored frequencies 0 to 600Hz during MFCC’s feature extraction, since addition of background noise causes phase shifts across (70 – 1000Hz).
* I ignored silence data in a signal to avoid misinterpretation of signal energy in a frame while extracting features(MFCC’s)

1. **What steps did you take to transform the original .wav files to prepare for feature extraction?**

Steps:

* convert raw PCM data to float audio range samples
* Silence removal in audio file
* Background noise suppression
* Divide complete signal in to 30 ms frames (for stationary behavior)

To maintain stationary nature of speech signal.

1. **How well do you think this audio reflects real-world conditions? What types of sounds might be harder to handle, and what edge cases would you prioritize detecting or filtering first?**

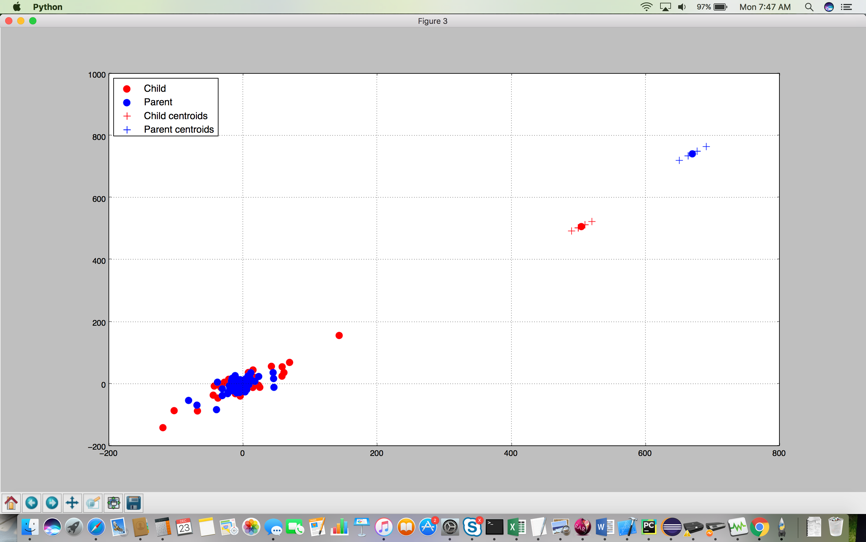
* Audio samples given has non stationary noise condition, this audio files will not cover complete real**-**world conditions.
* Background noise like multiple people talk at a time will be a challenging task. I prioritize background noise and double talk by using beam-forming and adaptive noise suppression techniques

1. **Is there anything you would want to add, optimize, or improve if you had more time?**

* I would have added more preprocessing units for background noise suppression and VAD algorithm to ignore silence in speech signal to make extracted features more informative.
* More data need to be collected and preprocessed for training ML models to make system robust.
* Optimize the code to reduce MIPS and help running it on low power device.

1. **If you were to annotate these audio files to train a supervised learning model, what labels would you include? Assume each row of the dataset represents a 10-second audio snippet, and provide an example column header.**

MFCC coefficients, LPC coefficients, signal length, SNR will be the main features to be considered to label features.



From MFCC coefficients Codebooks are generated using LBG algorithms for visibility about feature distribution. Centroids of Child and parent are generated from codebook.

From above observation there is a distinct in vocal tract from mother to child, This is evaluated by offline processing input data like silence removal, ignoring Background noise effecting frequencies

# Part 2: Modeling

## Task

Your challenge is to build a model to detect the gender of a voice. Always predicting "male" will achieve 50% accuracy. Our naive model based strictly on average dominant frequency achieves accuracy of 61% on training data and 59% on test data. We expect you can do significantly better.

1. **Determine which properties are statistically significant for determining gender. Report your results.**

🡪 All the given properties in input data are statistically significant, As it help to estimate person vocal information similar to MFCC, LPC feature extraction methods.

1. **Build a full logistic regression model and evaluate its performance. Report your results, including accuracy on the training and test set.**

Train Accuracy: 78%

Test Accuracy: 77.6%

1. **Train any model of your choice to achieve an accuracy above 80%.**

Using Tenforflow I could able to achieve below performance

Train Accuracy: 94%

Test Accuracy: 89%