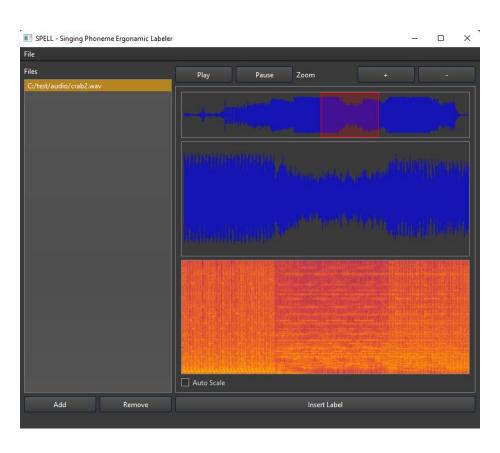
Singing Phoneme Ergonomic LabeLer - Milestone 3

Avinash Persaud, Nandith Narayan, Carlos Cepeda

Spectrogram

- Used the Fast Fourier Transform to create spectrograms.
- Added the spectrogram view to the main window.
- Multithreaded computation of spectrogram.
- Cached spectrogram of each sound file.
- Custom colormap support.

Spectrogram



Automated Phoneme Alignment

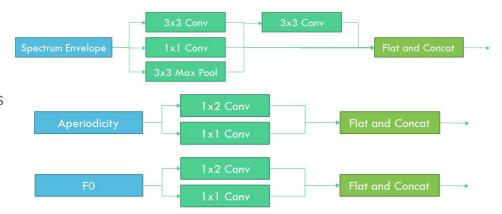
- Split into 2 main tasks
 - Phoneme boundary detection
 - Phoneme identification

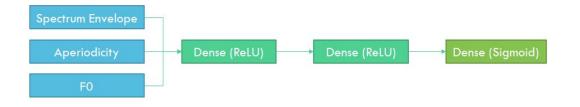
Phoneme Boundary Detection - Dataset

- 4 Datasets from the CMU Arctic project
- English Speech Data
- Used to make some voices for the Festival speech engine

Phoneme Boundary Detection - Model

- Inception-like CNN Model
- Outputs a transition probability
 - Peak detection is used to extract values



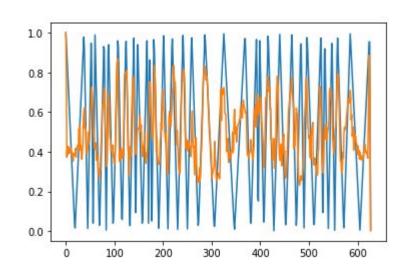


Phoneme Boundary Detection - Input Data

- First Revision
 - Vocoder Extracted Features
 - Spectral Envelope
 - Aperiodicity
 - FO
 - The 1st and 2nd derivatives as other "color channels"
 - 5 Frames
- Planned Next Step
 - 15 Mel Spectrogram + 1st and 2nd derivatives only

Phoneme Boundary Detection - Performance

- Numerical Accuracy was Poor
- Output is still semi-usable
- Noisy, but roughly fits target data
- Some tweaking necessary



Phoneme Identification

- Split the problem into chunks.
- Classify if a phoneme is a pause, consonant, or vowel.

Phoneme Identification - Dataset

- Initially used a dataset of children's songs.
- Couldn't separate phonemes.

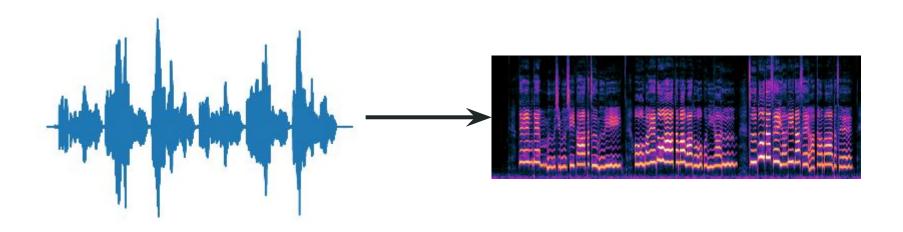
start end pitch syllable 2.4 2.85 61 ei 2.9813 3.5813 61 b_ii 3.5813 4.0687 68 s_ii 4.1813 4.6125 68 d_ii 5.4 5.25 70 ii 5.4 5.7938 70 e_f 6.5438 68 j_ii 7.1437 7.6125 66 ei_ch 7.7812 8.2125 66 ai 8.325 8.7938 65 j_ei 9.525 9.8625 63 e_l 4 9.8625 10.2 63 e_m 5 10.2 10.5 63 e_n 6 10.5 10.65 63 ou 7 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b	4	А	В	С	D
2.9813 3.5813 61 b_ii 3.5813 4.0687 68 s_ii 4.1813 4.6125 68 d_ii 5.4 5.7938 70 e_f 5.4 5.7938 70 e_f 6.5.9625 6.5438 68 j_ii 7.1437 7.6125 66 ei_ch 7.7812 8.2125 66 ai 8.325 8.7938 65 j_ei 8.9625 9.2625 65 k_ei 9.525 9.8625 63 e_l 9.525 9.8625 63 e_l 9.8625 10.2 63 e_m 10.5 10.5 63 ou 10.7625 11.175 61 p_ii 11.925 12.3938 68 k_y_uu 12.5625 12.975 68 a_r 11.925 12.3938 66 e_s 11.43812 15 65 t_ii 14.3812 15 65 t_ii 15.4313 65 y_uu 15.5813 15.9375 63 v_ii 16.8 17.0813 68 d_ao 17.0813 17.4 68 b_eo_l 17.0813 17.4 68 b_eo_l 17.0813 17.4 68 b_eo_l 17.9625 18.4688 66 e_k_s 19.1438 19.7812 65 w_ai 19.7812 20.3438 63 z_ii 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai		start	end	pitch	syllable
3.5813 4.0687 68 s_ii 4.1813 4.6125 68 d_ii 5.4.8 5.25 70 ii 5.4.5.7938 70 e_f 6.5.9625 6.5438 68 j_ii 7.1437 7.6125 66 ei_ch 7.7812 8.2125 66 ai 8.325 8.7938 65 j_ei 2.8.9625 9.2625 65 k_ei 3.9.525 9.8625 63 e_l 4.9.8625 10.2 63 e_m 6.10.5 10.65 63 ou 6.10.5 10.65 63 ou 6.10.5 10.65 63 ou 6.10.5 10.65 63 ou 6.10.5 12.3938 68 k_y_uu 6.10.7 12.3938 68 k_y_uu 6.10.8 13.1813 13.7438 66 e_s 6.10.8 13.1813 13.7438 66 e_s 6.10.9 13.1813 13.7438 66 e_s 6.10.9 13.1813 13.7438 66 e_s 6.10.9 15.4313 65 y_uu 6.10.9 15.4313 65 y_uu 6.10.9 15.4313 68 d_ao 6.10.9 15.4313 69 d_ao 6.		2.4	2.85	61	ei
4.1813		2.9813	3.5813	61	b_ii
4.8 5.25 70 ii 5.4 5.7938 70 e_f 5.9625 6.5438 68 j_ii 7.1437 7.6125 66 ei_ch 7.7812 8.2125 66 ai 8.325 8.7938 65 j_ei 2 8.9625 9.2625 65 k_ei 3 9.525 9.8625 63 e_l 4 9.8625 10.2 63 e_m 5 10.2 10.5 63 e_m 6 10.5 10.65 63 ou 7 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 12.5625 12.975 68 a_r 13.1813 13.7438 66 e_s 11.4.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 2 20.3438 20.7188 63 z_ii 2 1.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	ı	3.5813	4.0687	68	s_ii
5.4 5.7938 70 e_f 5.9625 6.5438 68 j_ii 7.1437 7.6125 66 ei_ch 7.7812 8.2125 66 ai 8.325 8.7938 65 j_ei 2 8.9625 9.2625 65 k_ei 3 9.525 9.8625 63 e_l 4 9.8625 10.2 63 e_m 5 10.2 10.5 63 e_m 6 10.5 10.65 63 ou 7 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 11.4.3812 15 65 t_ii 14.3812 15 65 t_ii 15.4313 65 y_uu 15.5813 15.9375 63 v_ii 16.8 17.0813 68 d_ao 17.0813 17.4 68 b_eo_l 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 20.3438 20.7188 63 z_ii 21.6 22.2 61 n_au 2 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai		4.1813	4.6125	68	d_ii
5.9625 6.5438 68 j_ii 7.1437 7.6125 66 ei_ch 7.7812 8.2125 66 ai 8.325 8.7938 65 j_ei 8.9625 9.2625 65 k_ei 9.8625 10.2 63 e_m 10.2 10.5 63 e_m 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 14.3812 15 65 t_ii 14.3812 15 65 t_ii 15.4313 66 e_s 11.43812 15 65 t_ii 15.4313 68 d_ao 17.0813 17.4 68 b_eo_l 17.0813 17.4 68 b_eo_l 17.0813 17.4 68 b_eo_l 17.9625 18.4688 66 e_k_s 19.1438 19.7812 65 w_ai 19.7812 20.3438 63 z_ii 11.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai		4.8	5.25	70	ii
7.1437 7.6125 66 ei_ch 7.7812 8.2125 66 ai 8.325 8.7938 65 j_ei 8.9625 9.2625 65 k_ei 9.525 9.8625 63 e_l 9.8625 10.2 63 e_m 10.2 10.5 63 e_n 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 11.31813 13.7438 66 e_s 11.4.3812 15 65 t_ii 15.4313 65 y_uu 15.5813 15.9375 63 v_ii 16.8 17.0813 68 d_ao 17.0813 17.4 68 b_eo_l 17.4 17.775 68 y_uu 17.9625 18.4688 66 e_k_s 19.1438 19.7812 65 w_ai 19.7812 20.3438 63 z_ii 11.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai		5.4	5.7938	70	e_f
0 7.7812 8.2125 66 ai 1 8.325 8.7938 65 j_ei 2 8.9625 9.2625 65 k_ei 3 9.525 9.8625 63 e_l 4 9.8625 10.2 63 e_m 5 10.2 10.5 63 e_n 6 10.5 10.65 63 ou 7 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 0 13.1813 13.7438 66 e_s 1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c		5.9625	6.5438	68	j_ii
1 8.325 8.7938 65 j_ei 2 8.9625 9.2625 65 k_ei 3 9.525 9.8625 63 e_l 4 9.8625 10.2 63 e_m 5 10.2 10.5 63 e_n 6 10.5 10.65 63 ou 7 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 0 13.1813 13.7438 66 e_s 1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 2 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai		7.1437	7.6125	66	ei_ch
2 8.9625 9.2625 65 k_ei 3 9.525 9.8625 63 e_l 4 9.8625 10.2 63 e_m 5 10.2 10.5 63 e_n 6 10.5 10.65 63 ou 7 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 0 13.1813 13.7438 66 e_s 1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	0	7.7812	8.2125	66	ai
9.525 9.8625 63 e_l 9.8625 10.2 63 e_m 10.2 10.5 63 e_n 10.5 10.65 63 ou 10.7625 11.175 61 p_ii 11.925 12.3938 68 k_y_uu 12.5625 12.975 68 a_r 13.1813 13.7438 66 e_s 1 14.3812 15 65 t_ii 15.4313 65 y_uu 15.5813 15.9375 63 v_ii 16.8 17.0813 68 d_ao 17.0813 17.4 68 b_eo_l 17.4 17.775 68 y_uu 17.9625 18.4688 66 e_k_s 19.1438 19.7812 65 w_ai 19.7812 20.3438 65 eo_n_c 20.3438 20.7188 63 z_ii 121.6 22.2 61 n_au 22.22.22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	1	8.325	8.7938	65	j_ei
4 9.8625 10.2 63 e_m 5 10.2 10.5 63 e_n 6 10.5 10.65 63 ou 7 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 0 13.1813 13.7438 66 e_s 1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 9 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	2	8.9625	9.2625	65	k_ei
5 10.2 10.5 63 e_n 6 10.5 10.65 63 ou 7 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 0 13.1813 13.7438 66 e_s 1 14.3812 15 65 t_iii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_iii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai <td>3</td> <td>9.525</td> <td>9.8625</td> <td>63</td> <td>e_l</td>	3	9.525	9.8625	63	e_l
6 10.5 10.65 63 ou 7 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 0 13.1813 13.7438 66 e_s 1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.7812 61 ai 3 22.7812 23.4 68 m_ai	4	9.8625	10.2	63	e_m
7 10.7625 11.175 61 p_ii 8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 0 13.1813 13.7438 66 e_s 1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	5	10.2	10.5	63	e_n
8 11.925 12.3938 68 k_y_uu 9 12.5625 12.975 68 a_r 0 13.1813 13.7438 66 e_s 1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	6	10.5	10.65	63	ou
9 12.5625 12.975 68 a_r 0 13.1813 13.7438 66 e_s 1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	7	10.7625	11.175	61	p_ii
13.1813 13.7438 66 e_s 1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	8	11.925	12.3938	68	k_y_uu
1 14.3812 15 65 t_ii 2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	9	12.5625	12.975	68	a_r
2 15 15.4313 65 y_uu 3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_o 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	0	13.1813	13.7438	66	e_s
3 15.5813 15.9375 63 v_ii 4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	1	14.3812	15	65	t_ii
4 16.8 17.0813 68 d_ao 5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	2	15	15.4313	65	y_uu
5 17.0813 17.4 68 b_eo_l 6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	3	15.5813	15.9375	63	v_ii
6 17.4 17.775 68 y_uu 7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	4	16.8	17.0813	68	d_ao
7 17.9625 18.4688 66 e_k_s 8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	5	17.0813	17.4	68	b_eo_l
8 19.1438 19.7812 65 w_ai 9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	6	17.4	17.775	68	y_uu
9 19.7812 20.3438 65 eo_n_c 0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	7	17.9625	18.4688	66	e_k_s
0 20.3438 20.7188 63 z_ii 1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	8	19.1438	19.7812	65	w_ai
1 21.6 22.2 61 n_au 2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	9	19.7812	20.3438	65	eo_n_c
2 22.2 22.7812 61 ai 3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	0	20.3438	20.7188	63	z_ii
3 22.7812 23.4 68 n_ou 4 23.4 23.8688 68 m_ai	-	21.6	22.2	61	n_au
4 23.4 23.8688 68 m_ai		22.2	22.7812	61	ai
	3		23.4		
5 24 24.45 70 ei					
	5	24	24.45	70	ei

Phoneme Identification - Dataset

- Nagoya Institute of Technology 70 song dataset
 - Only 31 publicly available
- Professional Japanese Female Singer
- Traditional Japanese Folk Songs
 - From what we can identify
- Distributed in the HTS Japanese Song demo
 - The full dataset is used in the Sinsy synthesizer system under the voice f00001j Yoko

Phoneme Identification - Data Preprocessing

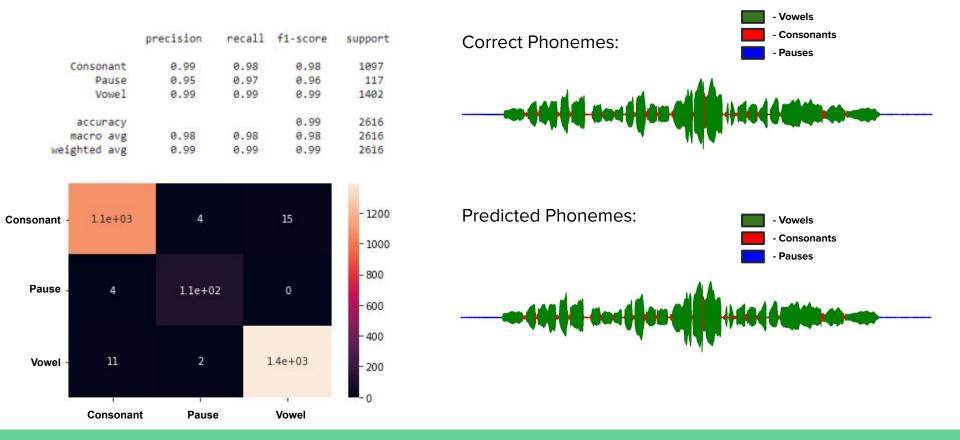
Created MEL spectrogram from sound.



Phoneme Identification - Model

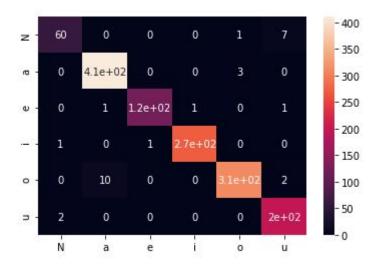
- Simple CNN, with 5 convolutional layers
- Input is 20 by 40 log of the MEL spectrogram
- Output is one hot vector representing if a phoneme is a pause, consonant, or vowel.

Phoneme Identification - Results on Testset



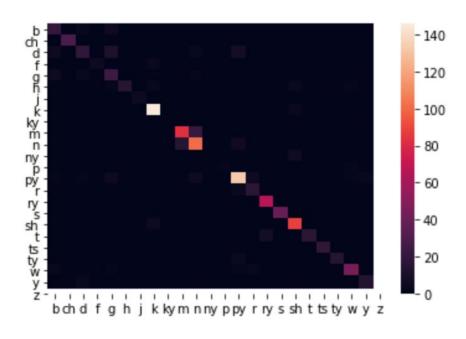
Phoneme Identification - Vowel Phoneme Results

	precision	recall	f1-score	support	
N	0.95	0.88	0.92	68	
а	0.97	0.99	0.98	413	
е	0.99	0.98	0.98	126	
i	1.00	0.99	0.99	274	
0	0.99	0.96	0.97	323	
u	0.95	0.99	0.97	199	
accuracy			0.98	1403	
macro avg	0.98	0.97	0.97	1403	
weighted avg	0.98	0.98	0.98	1403	



Phoneme Identification - Consonant Phoneme Results

		precision	recall	f1-score	support
	b	0.53	0.62	0.57	32
	ch	1.00	0.97	0.98	30
	d	0.59	0.42	0.49	45
	f	0.71	0.42	0.53	12
	g	0.46	0.62	0.53	39
	h	0.83	0.52	0.64	29
	j	1.00	0.62	0.77	8
	k	0.92	0.96	0.94	152
	ky	1.00	0.33	0.50	3
	m	0.81	0.78	0.79	106
	n	0.76	0.81	0.78	125
	р	1.00	0.14	0.25	7
	ру	1.00	0.40	0.57	5
	r	0.85	0.85	0.85	157
	ry	0.62	0.70	0.65	23
	5	0.90	1.00	0.95	66
	sh	0.97	1.00	0.99	39
	t	0.82	0.95	0.88	93
	ts	1.00	0.71	0.83	24
	ty	0.95	1.00	0.97	19
	W	0.94	0.71	0.81	21
	У	0.87	0.82	0.84	55
	Z	0.67	0.74	0.70	19
accur	acy			0.81	1109
macro	avg	0.83	0.70	0.73	1109
weighted	avg	0.82	0.81	0.81	1109



Questions?