

**ALTERNATE PRONUNCIATION
&
REGARDING ISSUES**

**Presented by
Muhammad Ayub**



Center for Language Engineering (CLE)
Al-Khwarizmi Institute of Computer Science
University of Engineering and Technology
Lahore, Pakistan

INTRODUCTION

- Pakistan is a multilingual country as almost 59 different languages are being spoken.
- The names of 139 districts of Pakistan are brought under the influence of six major accents i.e. Urdu, Punjabi, Sindhi, Balochi, Pashto and Saraiki against their standard pronunciation to analyzed the changes.
- A variation in the standard pronunciation is taken into account to go through some specific measures to allocate for the Alternate Pronunciation or it is removed.
- Here a new concept of variation in the standard pronunciation is pondered upon which I will discuss in the subsequent slides.

ALTERNATE PRONUNCIATION

- Definition
- Criteria of AP



Alternate Pronunciation & regarding issues

sp00256_z057_pun_M_dt008_ver01.wav AP VDM D_ZA_AFRA_ABA_AD_D

sp00293_z057_pun_M_dt008_ver01.wav AP VDM D_ZA_AFRA_ABA_AD_D

sp00334_z057_pun_M_dt008_ver01.wav

sp00410_z072_pun_F_dt008_ver01.wav AP VDM D_ZA_AFRA_ABA_AD_D

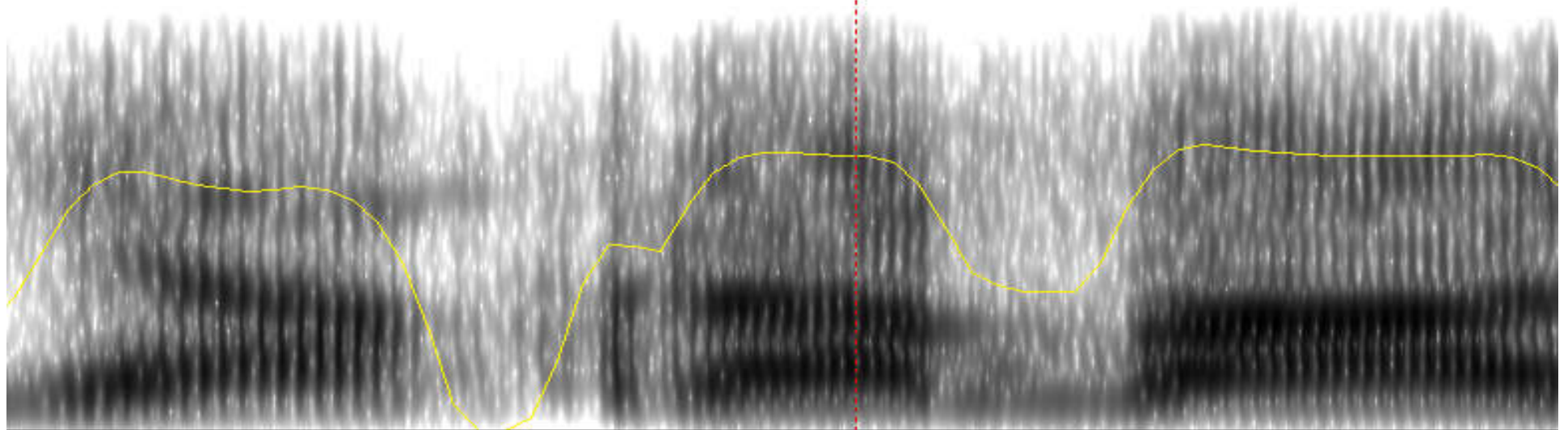
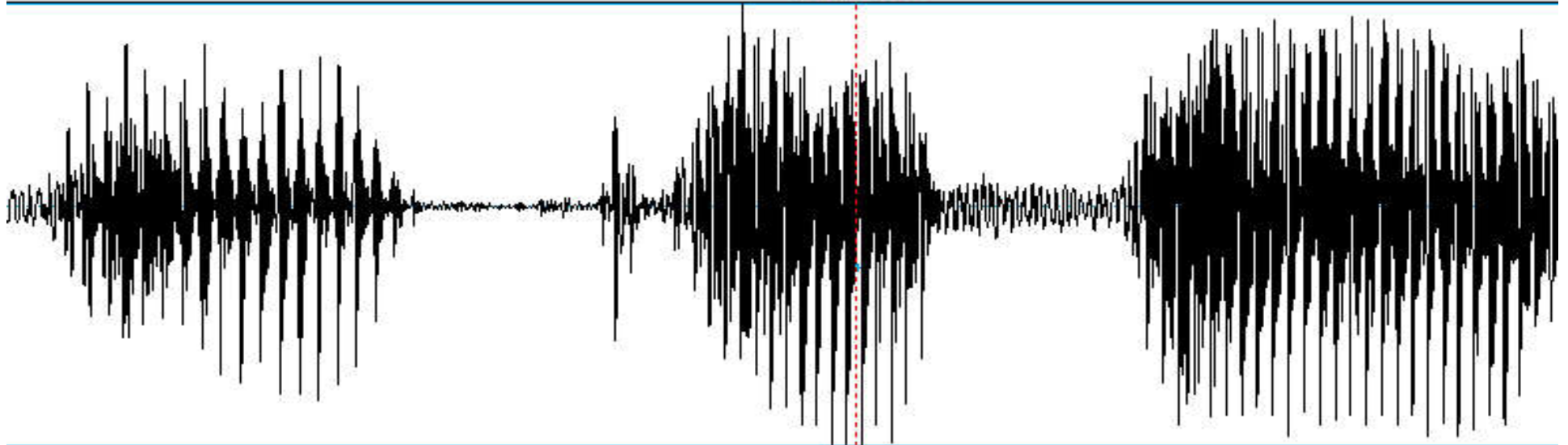
sp00439_z140_pun_F_dt008_ver01.wav AP VDM D_ZA_AFRA_ABA_AD_D

sp00453_z079_pun_F_dt008_ver01.wav AP VDM D_ZA_AFRA_ABA_AD_D



Pitch Intensity Formant Pulses

0.475063



0.475063

0.475063

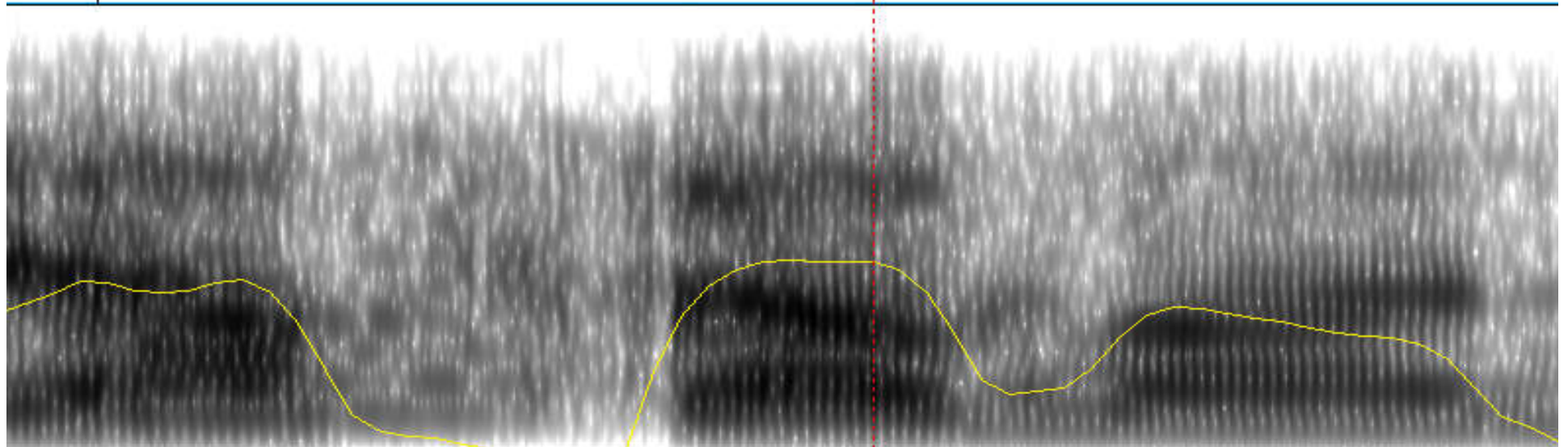
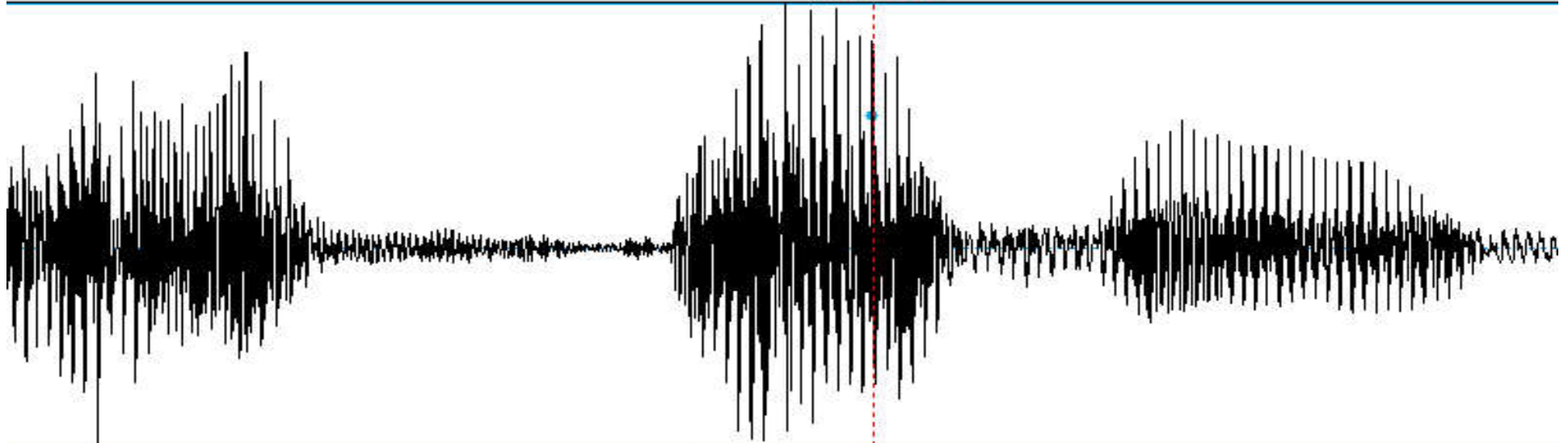
Visible part 0.950125 seconds

Total duration 0.950125 seconds

l

m Pitch Intensity Formant Pulses

0.442563



0.442563

0.442563

Visible part 0.885125 seconds

Total duration 0.885125 seconds



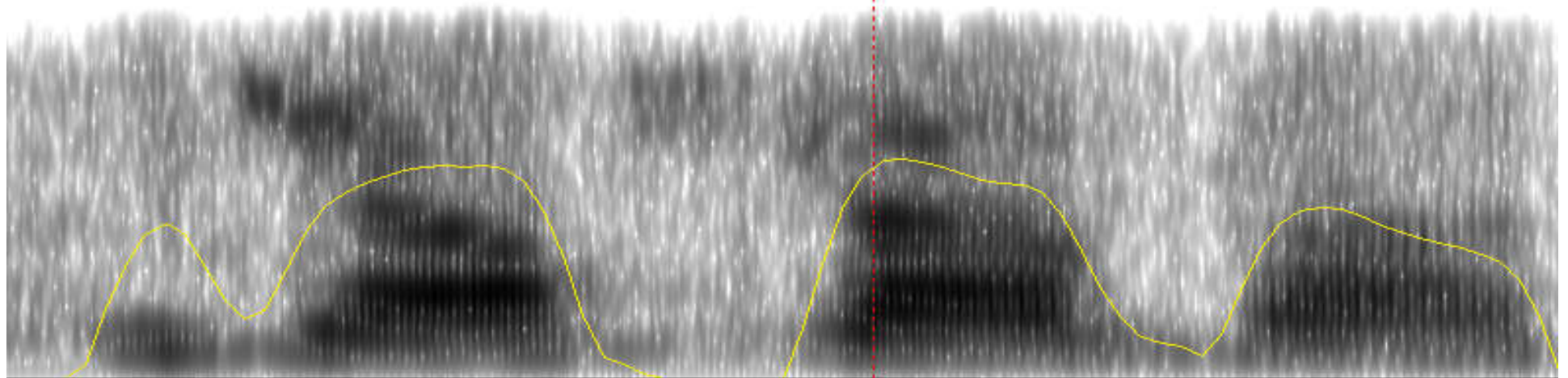
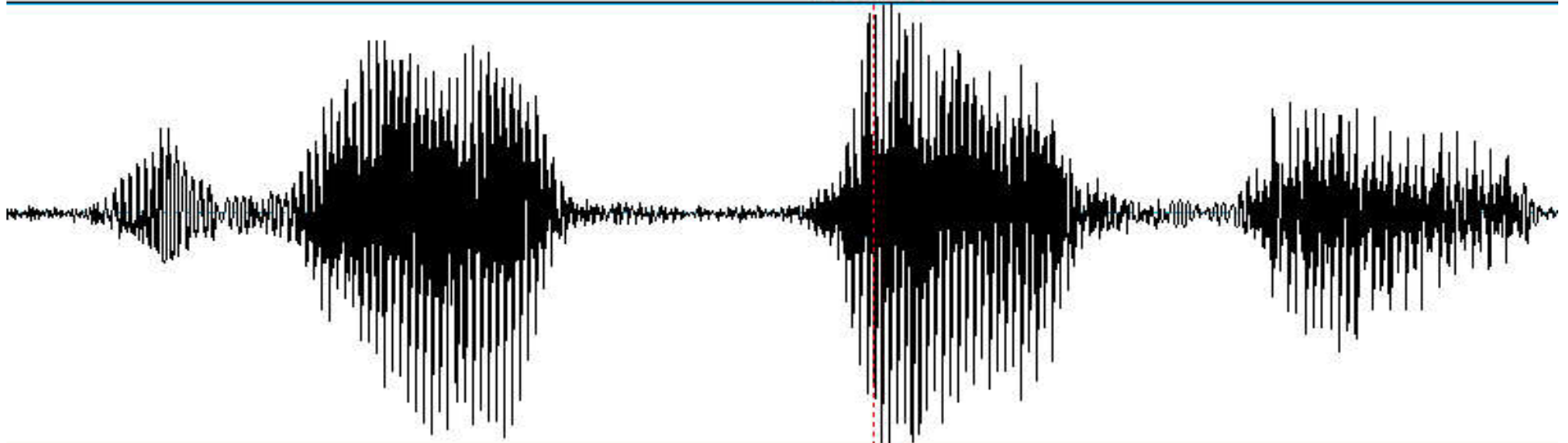
Alternate Pronunciation & regarding issues

sp01754_025_urd_F_dt008_ver01.wav AP	VDM	D_ZA_AFRA_AB_AD_D
sp01957_025_urd_F_dt008_ver01.wav AP	VDM	D_ZA_AFRA_AB_AD_D
sp01971_025_urd_F_dt008_ver01.wav AP	VDM	D_ZA_AFRA_ABA_AD_D
sp02021_025_urd_F_dt008_ver01.wav		
sp02099_025_urd_F_dt008_ver01.wav AP	VDM	D_ZA_AFRA_ABA_AD_D
sp02168_025_urd_F_dt008_ver01.wav AP	VDM	D_ZA_AFRA_ABA_AD_D



m Pitch Intensity Formant Pulses

0.607688



0.607688

0.607688

Visible part 1.215375 seconds

Total duration 1.215375 seconds



Alternate Pronunciation & regarding issues

sp01391_z025_pus_F_dt008_ver01.wav AP

CSP/M

D_ZA_AFARA_AVA_AD_D

sp01396_z024_pus_F_dt008_ver01.wav AP

VDM

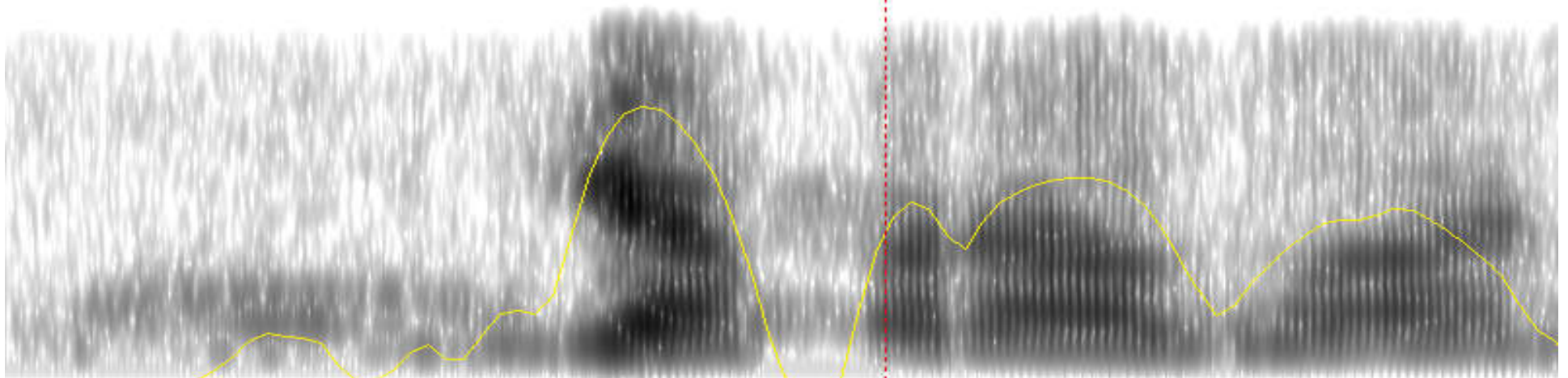
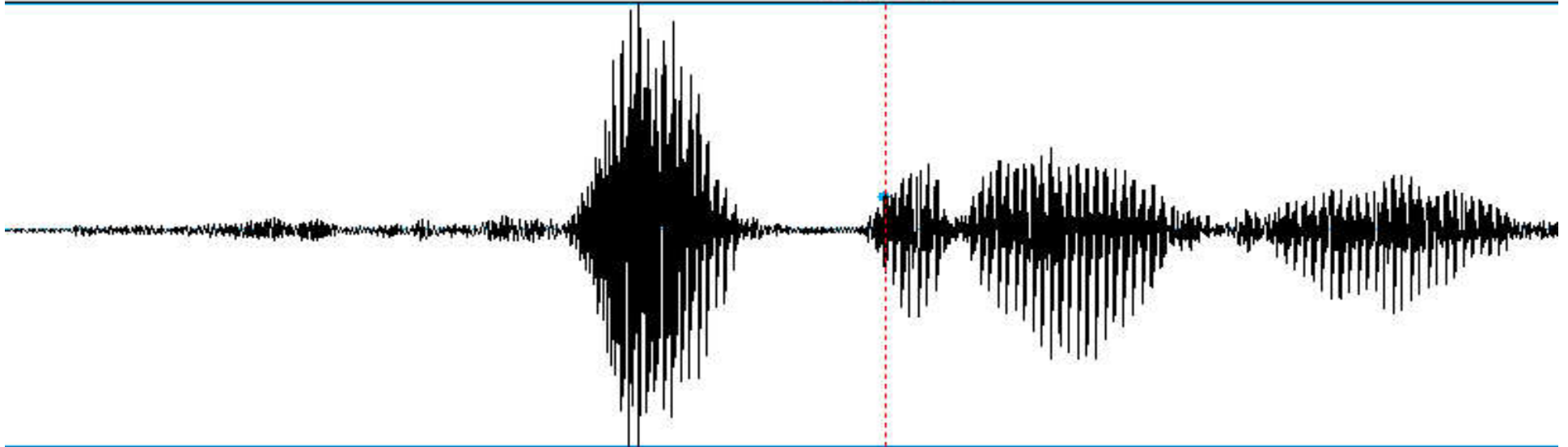
D_ZA_AFRA_ABA_AD_D



r01

rum Pitch Intensity Formant Pulses

0.670312



0.670312

0.670312

Visible part 1.340625 seconds

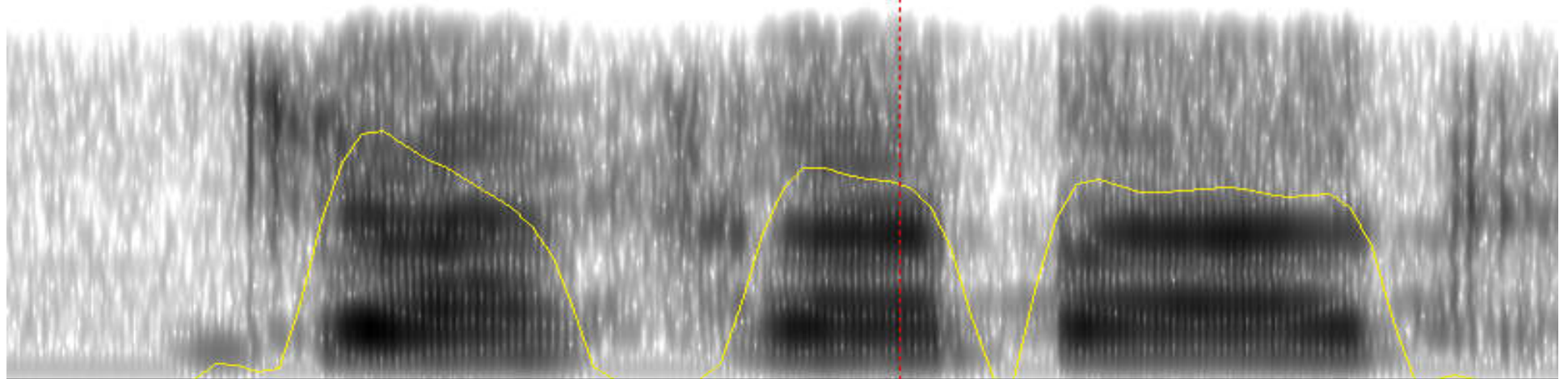
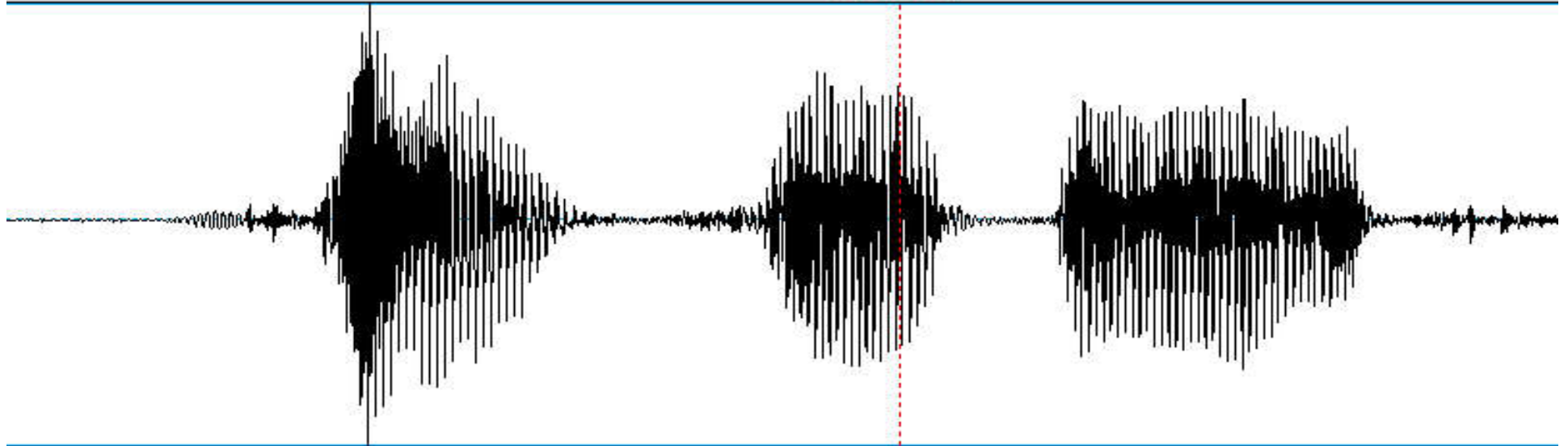
Total duration 1.340625 seconds



ver01

Spectrum Pitch Intensity Formant Pulses

0.565000



0.565000

0.565000

Visible part 1.130000 seconds

Total duration 1.130000 seconds



Alternate Pronunciation & regarding issues

sp01392_z025_bal_F_dt008_ver01.wav AP VDM D_ZA_AFRA_ABA_AD_D

sp01456_z014_bra_F_dt008_ver01.wav AP VDM D_ZA_AFRA_ABA_AD_D

sp01466_z011_bal_F_dt008_ver01.wav AP VDM D_ZA_AFRA_ABA_AD_D

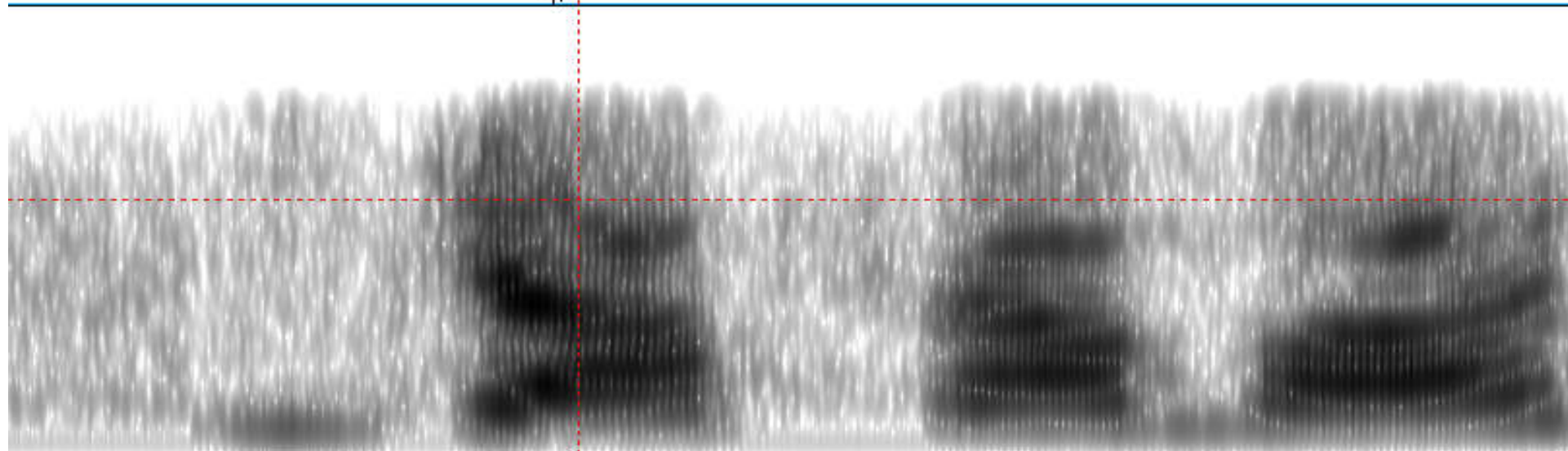
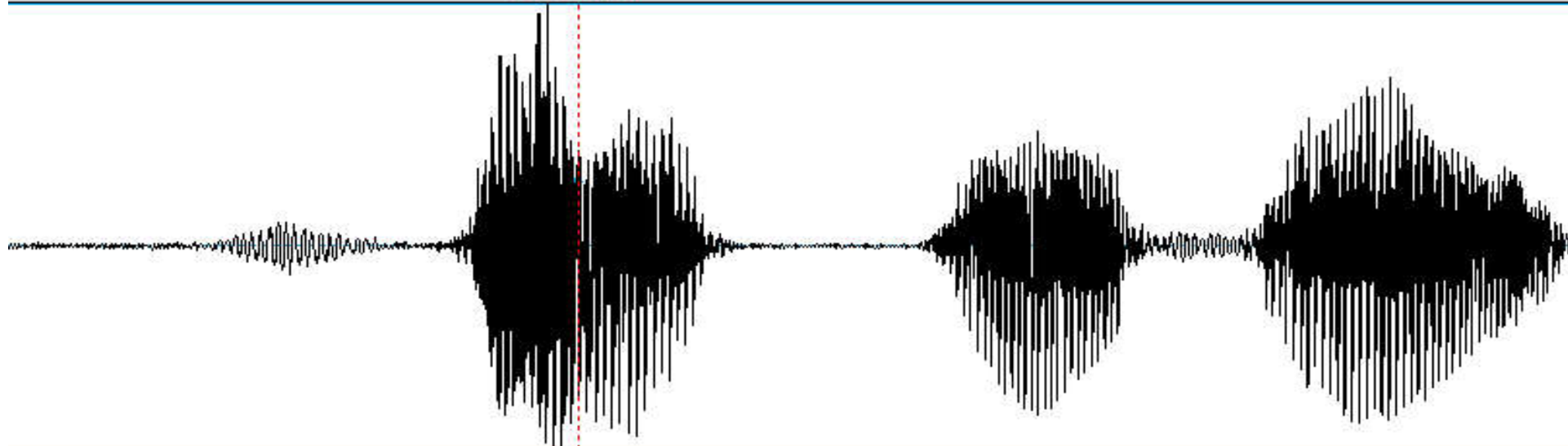
sp02679_z140_bal_M_dt008_ver01.wav AP VDM D_ZA_AFRA_ABA_AD_D



ver01

Spectrum Pitch Intensity Formant Pulses

0.405871



0.405871

0.757254

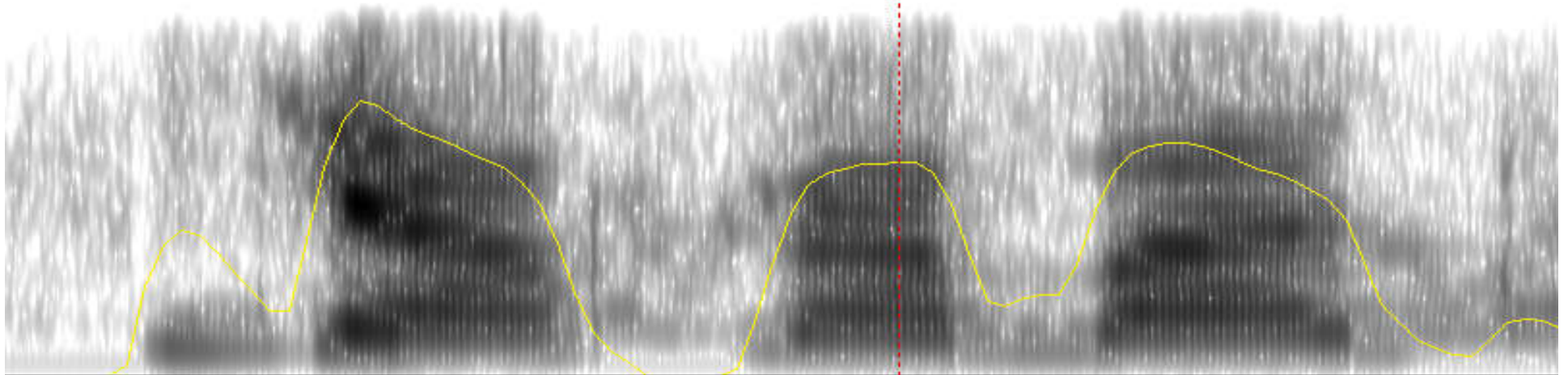
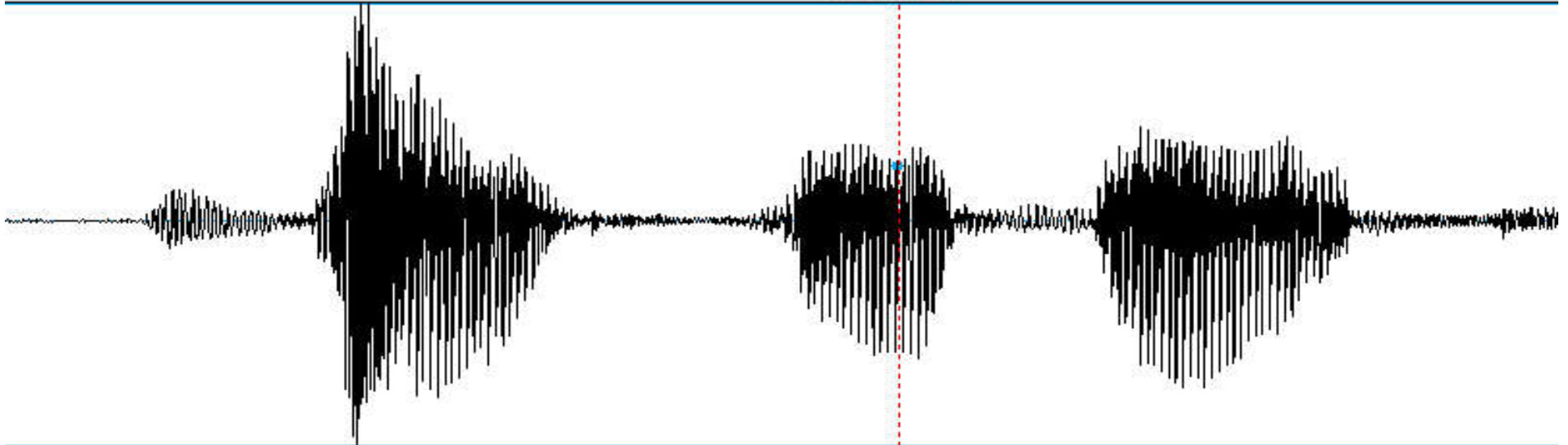
Visible part 1.163125 seconds

Total duration 1.163125 seconds

er01

Spectrum Pitch Intensity Formant Pulses

0.663500



0.663500

0.663500

Visible part 1.327000 seconds

Total duration 1.327000 seconds



Alternate Pronunciation & regarding issues

DATA ANALYSIS OF D_ZA_AFARA_ABA_AD_D

- Clean files of Punjabi speakers =40
- No. of AP =19
- AP =47.5%
- Clean files of Urdu speakers =23
- No. of AP =13
- AP =62 %
- Clean files of Balochi speakers = 9
- No. of AP =4
- AP =44.4 %
- Clean files of Pashto speakers =53
- No. of AP =15
- AP = 28.3 %



READ ME.TEXT

- D_ZA_AFARA_ABA_AD_D:
- This district folder contains two types of AP (alternate pronunciation). These AP's are marked as AP1 and AP2. Their respective transcriptions are given below:
- AP 1 :D_ZA_AFRA_ABA_AD_D
- AP 2 :D_ZA_AFARA_AVA_AD_D

SAME TRANSCRIPTION

- But Pronunciation is different
- D_ZA_AFAR_[PAU]A_ABA_AD_D
- D_ZA_AFARA_A_[PAU]BA_AD_D

HENCE ALTERNATE PRONUNCIATION

- Definition:
- Alternate Pronunciation is a variation of standard pronunciation in which substitution, deletion, insertion of vowel and substitution of consonant is analyzed if no. of instances show a general trend towards that accent.

Alternate Pronunciation & regarding issues

sp00436_z140_pun_F_dt021_ver01.wav AP VS NA_AS^IRA_ABA_AD_D

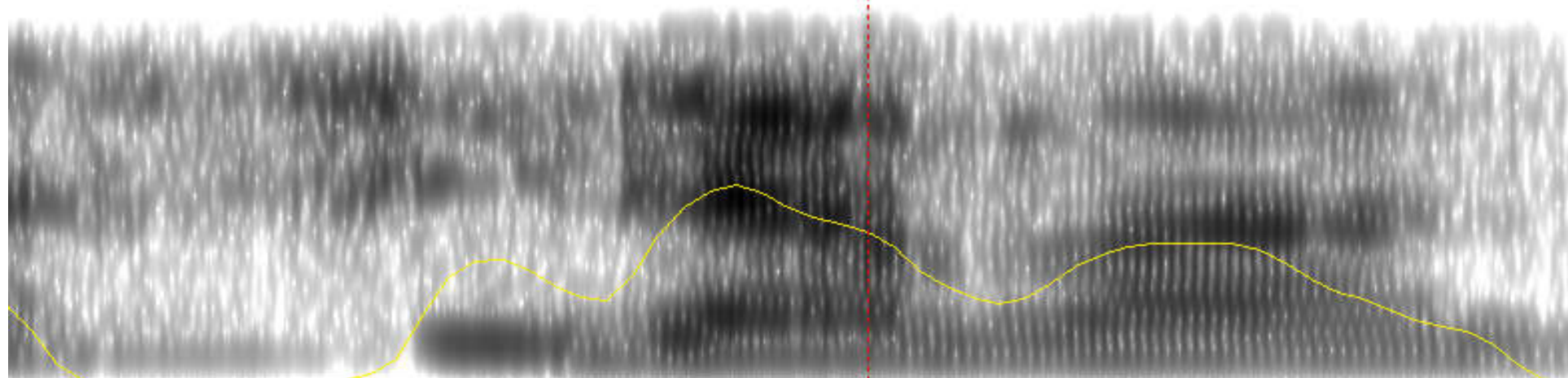
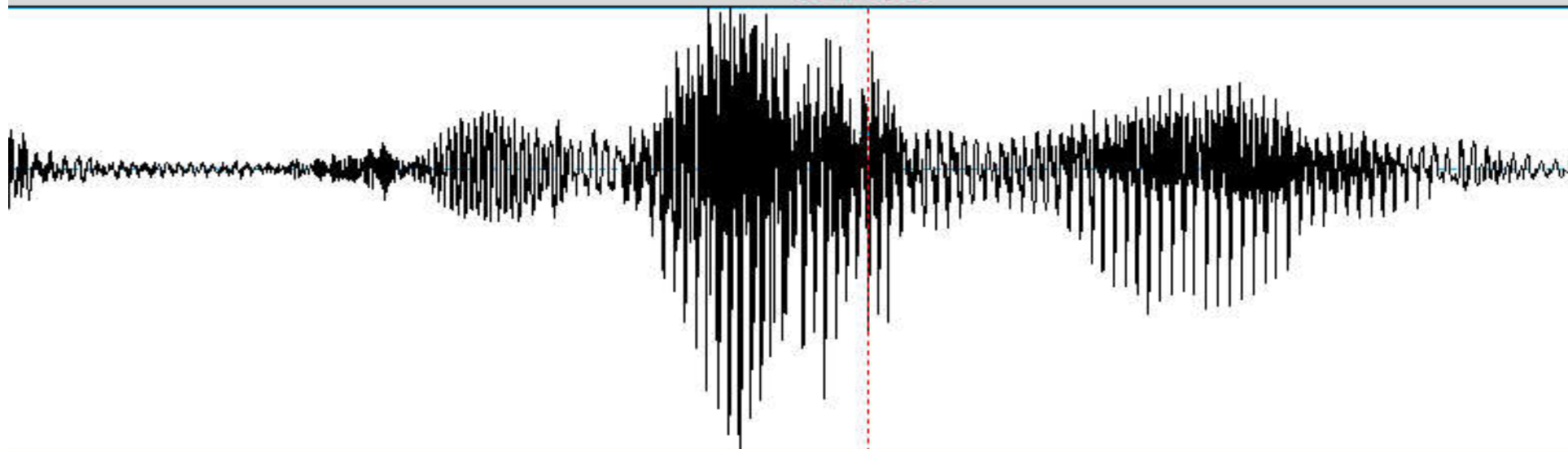
sp00452_z088_pun_F_dt021_ver01.wav AP VS NA_AS^IRA_ABA_AD_D

sp00484_z057_pun_M_dt021_ver01.wav AP VS NA_AS^ARA_ABA_AD_D

sp00494_z057_pun_M_dt021_ver01.wav AP VS NA_AS^ARA_ABA_AD_D

Pitch Intensity Formant Pulses

0.471500



0.471500

0.471500

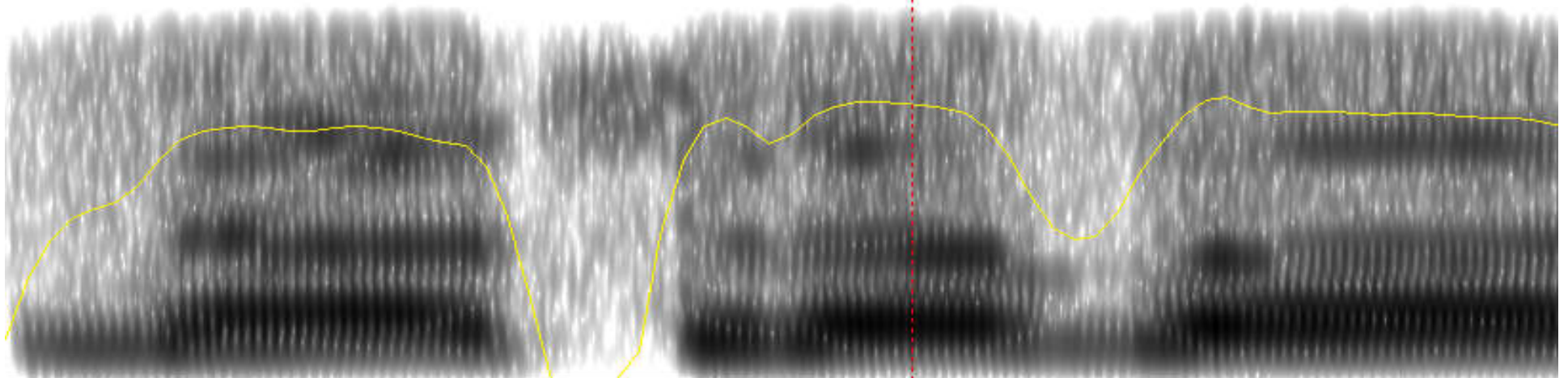
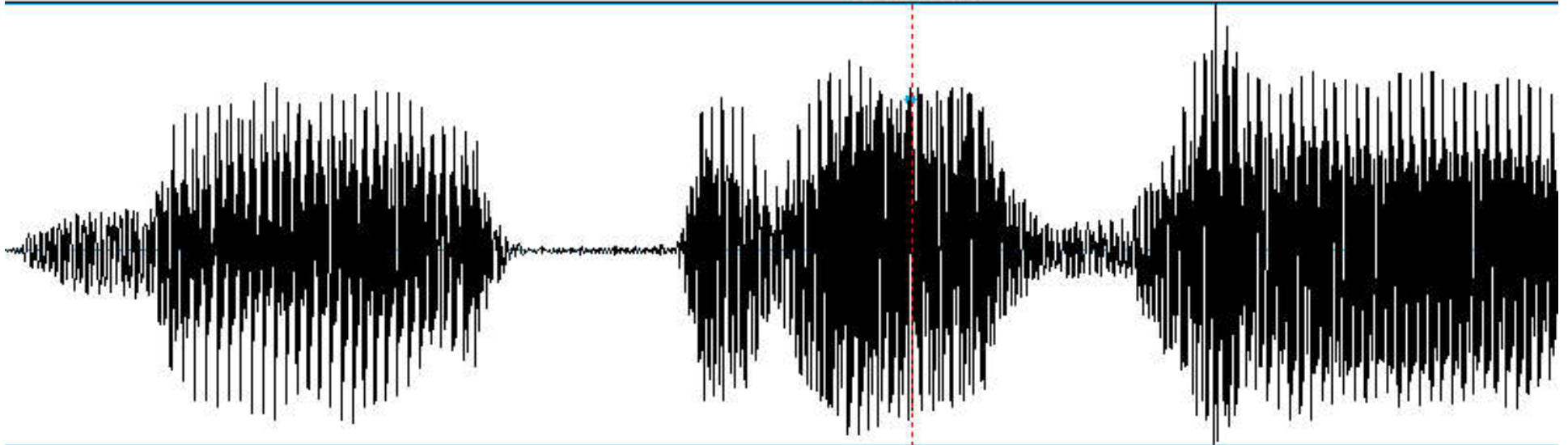
Visible part 0.943000 seconds

Total duration 0.943000 seconds

21_ver01

pectrum Pitch Intensity Formant Pulses

0.540000



0.540000

0.540000

Visible part 1.080000 seconds

Total duration 1.080000 seconds

Alternate Pronunciation & regarding issues

DATA ANALYSIS OF NASI_IRA_ABA_AD_D

- Clean files of Punjabi speakers = 31
- No. of AP = 18
- AP = 58 %
- Clean files of Urdu speakers = 21
- No. of AP = No

- Clean files of Balochi speakers = 6
- No. of AP = No

- Clean files of Pashto speakers = 52
- No. of AP = No
- No of RM = 4
- I f AP(suppose) =8 %



Alternate Pronunciation & regarding issues

sp01174_z044_pus_M_dt021_ver01.wav IP NASI_IRA_AVA_AD_D

sp01181_z044_pus_M_dt021_ver01.wav

sp01184_z044_pus_M_dt021_ver01.wav

sp01196_z044_pus_M_dt021_ver01.wav IP NASI_IRA_AVA_AD_D

sp01198_z044_pus_M_dt021_ver01.wav

sp01249_z045_pus_F_dt021_ver01.wav IP NASI_IRA_AVA_AD_D

sp01657_z052_pus_M_dt021_ver01.wav IP NASI_IRA_AVA_AD_D

Similarly

sp00991_z037_pus_M_dt021_ver01.wav RM VSD NA_ASIRA_ABA_AD_D

Alternate Pronunciation & regarding issues

sp01174_z044_pus_M_dt021_ver01.wav RM CSD NASI_IRA_AVA_AD_D

sp01181_z044_pus_M_dt021_ver01.wav

sp01184_z044_pus_M_dt021_ver01.wav

sp01196_z044_pus_M_dt021_ver01.wav RM CSD NASI_IRA_AVA_AD_D

sp01198_z044_pus_M_dt021_ver01.wav

sp01249_z045_pus_F_dt021_ver01.wav RM CSD NASI_IRA_AVA_AD_D

sp01657_z052_pus_M_dt021_ver01.wav RM CSD NASI_IRA_AVA_AD_D

Similarly

sp00991_z037_pus_M_dt021_ver01.wav RM VSD NA_ASIRA_ABA_AD_D



CONCLUSION

- Generally it is supposed that every variation in standard pronunciation is either discarded or marked as AP after making go through some specific parameters. In this research work, it has been proved that the variation in standard pronunciation will be processed as correct if it keeps the transcription unchanged i.e. it will neither be discarded nor marked as AP.

SUGGESTIONS

- Adjustment of AP according to no. of files
- Concept of a New Keyboard





	£ 3	\$ 4	% € 5	& 6	/ 7	(8) 9	= 0	? 1
W	E € A_E	R	T T_D	Y A_Y	U U_U	I I_I	O O_O	P P_H	
S T_S	D D_D	F	G G_G	H	J	K K_H	L	ç ò	
!)_Z	X	C	V	B	N N_G	M	; ,	: .	
Alt								Alt Gr	

CHANGE OF TRANSCRIPTION

- The transcription of code(103) MI_IRPU_URKHAS has been changed into MI_IRPU_URXA_AS .
- The transcription of code(121) JANUBI WAZIRISTAN has been changed into D_ZANU_UBI_IVAZI_IRIST_DA_AN
- The transcription of code(135) DIA_AMAR has been changed into D_DIA_AMAR
- The transcription of code(240) DO_OPA_E_HHAR has been changed into D_DO_OPA_E_HHAR

THANKS

by

Muhammad Ayub

مرکز تحقیقات لسانیات



Center for Language Engineering (CLE),
Al-Khwarizmi Institute of Computer Science,
University of Engineering and Technology
Lahore, Pakistan