

# DE CE NU PLACE VOCEA SINTETIZATĂ?

– CÂTEVA ELEMENTE DE  
COMPARAȚIE CU VOCEA UMANĂ

Horia-Nicolai Teodorescu\*,\*\*

Monica Feraru \*,\*\*

\* Romanian Academy, Institute for Computer Science, Iași, Romania

\*\* CERFS Center for Research, Technical University “Gh. Asachi”, Iași, Romania

# Structura

- Introducere
- Metodologie
- Rezultatele analizei
- Discuții
- Concluzii
- Direcții viitoare

# Introducere

- inteligibilitatea este dată doar de formanții inferiori (F1, F2)
- calitatea pronunției, nuanțele și bogăția vocii sunt date în mare măsură de formanții superiori
- o scurtă analiză a diferențelor care apar, la nivel de formanți, între vocea sintetizată și vocea naturală
- comparația privește singurul sintetizor comercial autohton pentru limba română

# Metodologie

- metoda de comparație privește două aspecte la nivel strict *formantic static*:

- valorile frecvențelor formanților

- valorile raporturilor frecvențelor formanților raportate la frecvența fundamentalei pentru acea vocală

- s-a folosit o singură voce umană, din corpusul adnotat SRoL, aleasă pe criteriul similarității valorii frecvenței fundamentale

# Detalii

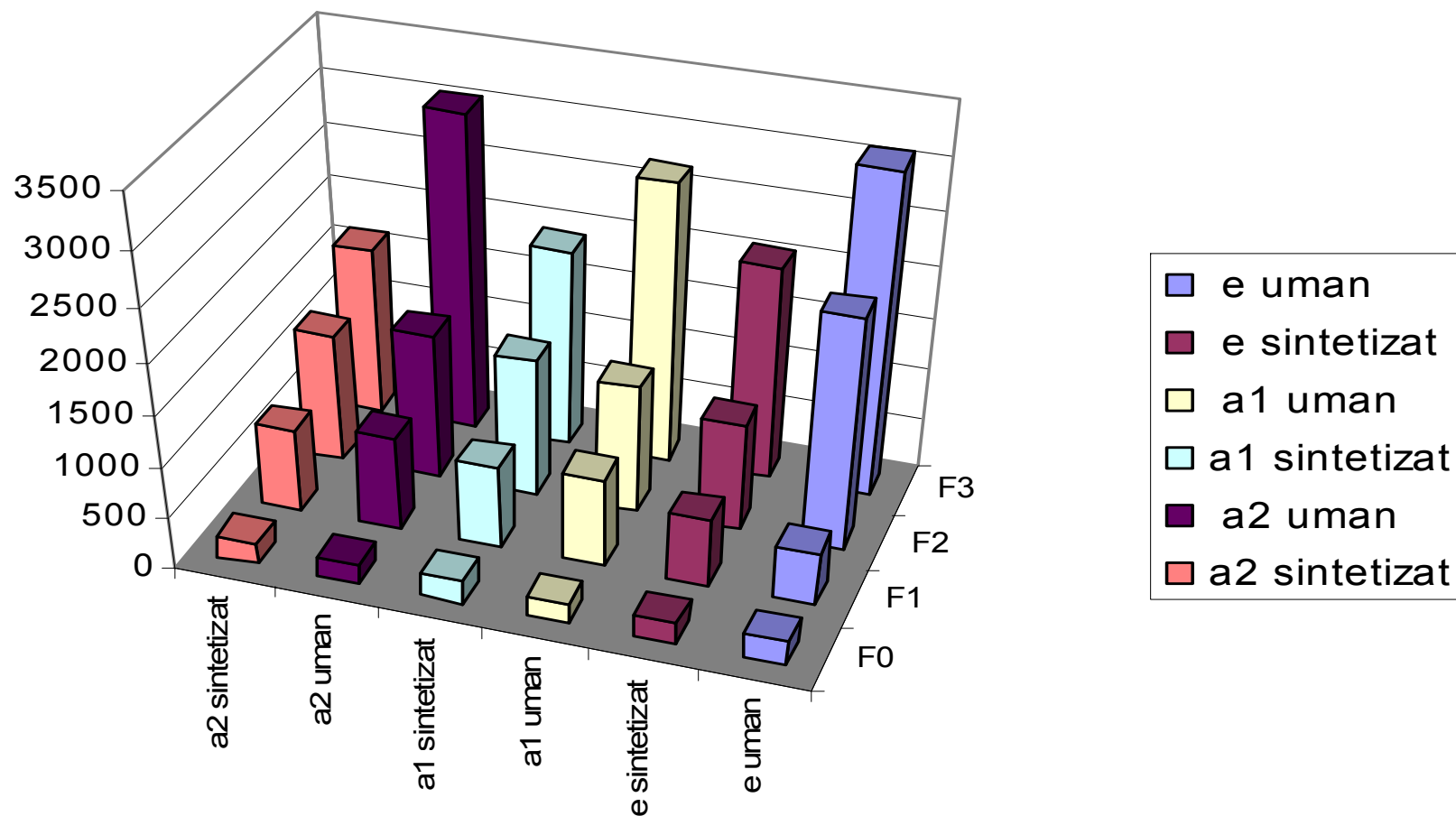
- vocea este feminină; 55555f - un vorbitor cu educație, fără patologie, voce regională din zona Moldovei, fără accent dialectal evident etc.
- sintetizorul discutat este unul de tip concatenativ; pentru sintetizor sunt folosite setările *default*, Viteza: 60, Intonația: 60, Format: .WAV - nu sunt date de către autorii
- s-a folosit utilitarul Praat™; cuvintele au fost segmentate manual pe foneme, am determinat valorile medii pe vocale ale frecvențelor fonemelor.
- cuvântul analizat este Aseară, iar propozițiile comparate sunt “Vine mama” și “Cine a făcut asta?”.

# Rezultatele analizei / “Vine mama” (I)

	i	i	e	e	a1	a1	a2	a2
	uman	sintetic	uman	sintetic	uman	sintetic	uman	sintetic
<b>F0</b>	222	234	242	226	196	225	200	218
<b>F1</b>	348	463	518	661	873	816	921	807
<b>F2</b>	<b>2690</b>	<b>864</b>	<b>2242</b>	<b>1054</b>	1255	1348	1427	1255
<b>F3</b>	<b>3792</b>	<b>2450</b>	3116	2083	2718	1928	3070	1642
<b>F1/F0</b>	1.57	1.98	2.14	2.92	4.45	3.62	4.60	3.70
<b>F2/F2</b>	12.11	3.69	9.26	4.66	6.40	5.99	7.13	5.75
<b>F3/F0</b>	<b>17.08</b>	<b>10.47</b>	12.87	9.21	<b>13.87</b>	<b>8.56</b>	15.35	7.53

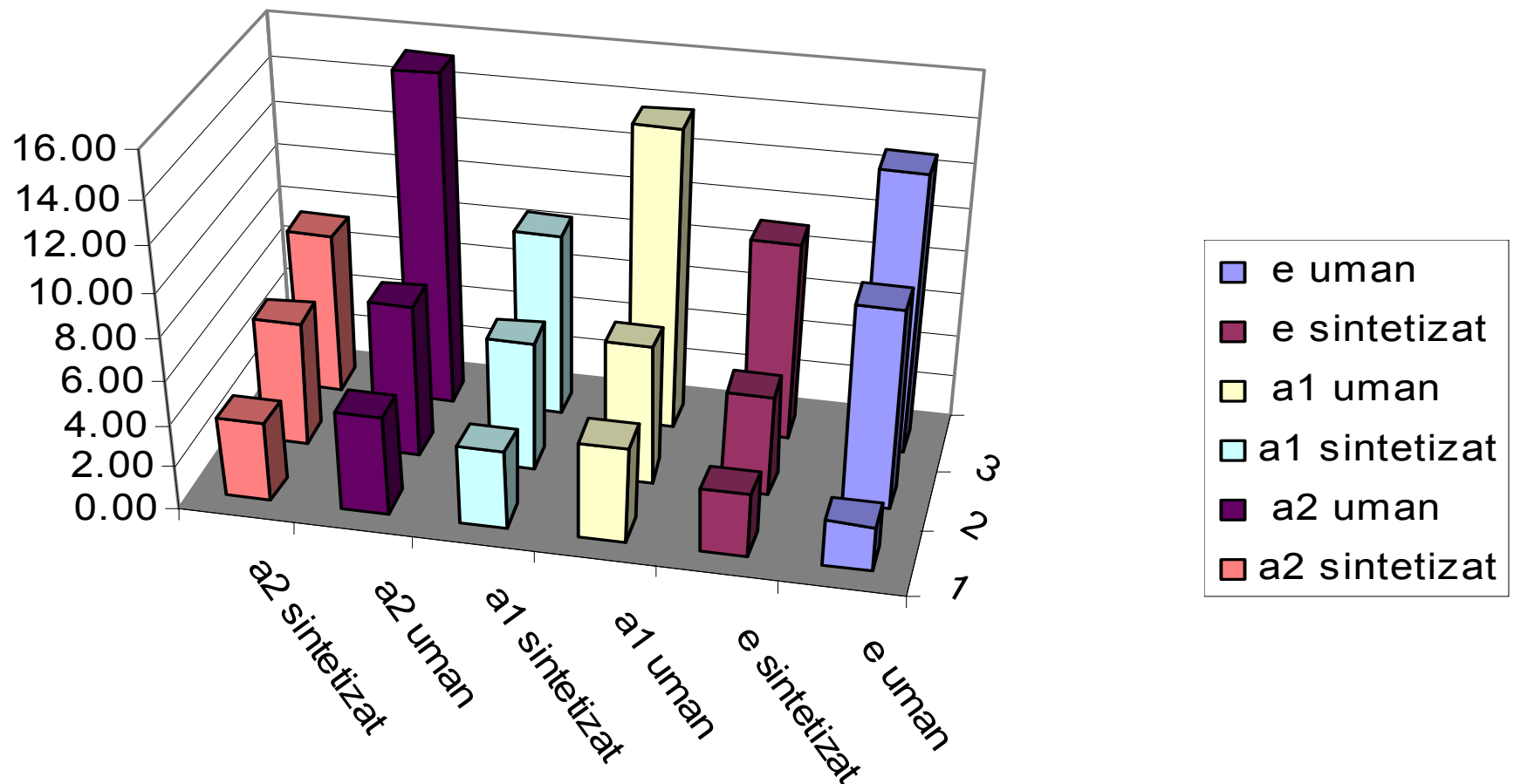
# Rezultatele analizei / “Vine mama” (II)

Vine mama - valori absolute F0, F1, F2 si F3



# Rezultatele analizei / “Vine mama” (III)

Vine mama - raporturi F1/F0, F2/F0 si F3/F0



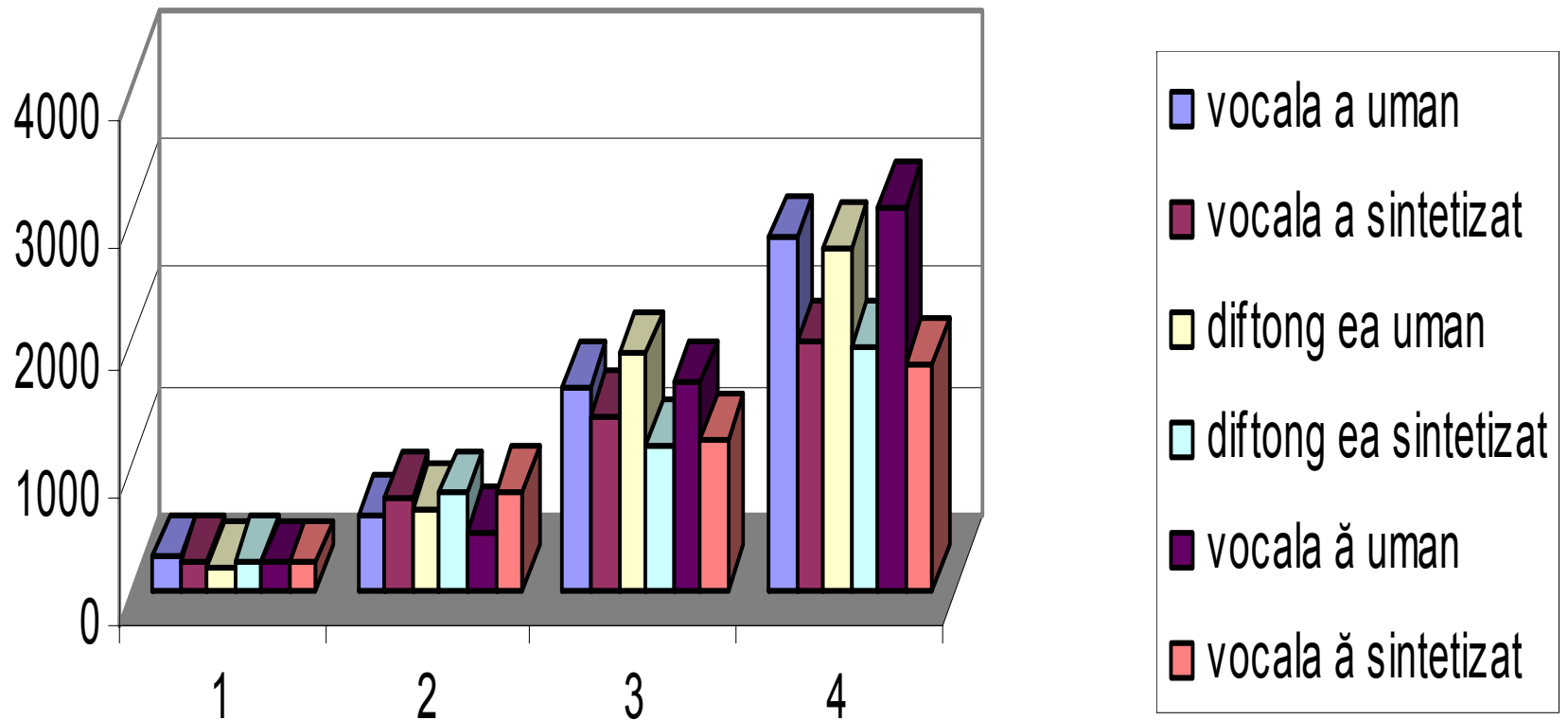


# Rezultatele analizei / “Aseară” (I)

	a	a	ea	ea	ă	ă
	uman	sintetic	uman	sintetic	uman	sintetic
<b>F0</b>	254	226	200	228	207	219
<b>F1</b>	593	752	652	769	<b>456</b>	<b>785</b>
<b>F2</b>	1634	1403	<b>1892</b>	<b>1176</b>	<b>1658</b>	<b>1212</b>
<b>F3</b>	<b>2806</b>	<b>1978</b>	<b>2736</b>	<b>1963</b>	3060	1827
<b>F1/F0</b>	<b>2.33</b>	<b>3.33</b>	3.26	3.37	2.20	3.58
<b>F2/F2</b>	6.43	6.21	9.46	5.16	8.01	5.53
<b>F3/F0</b>	11.05	8.75	13.68	8.61	14.78	8.34

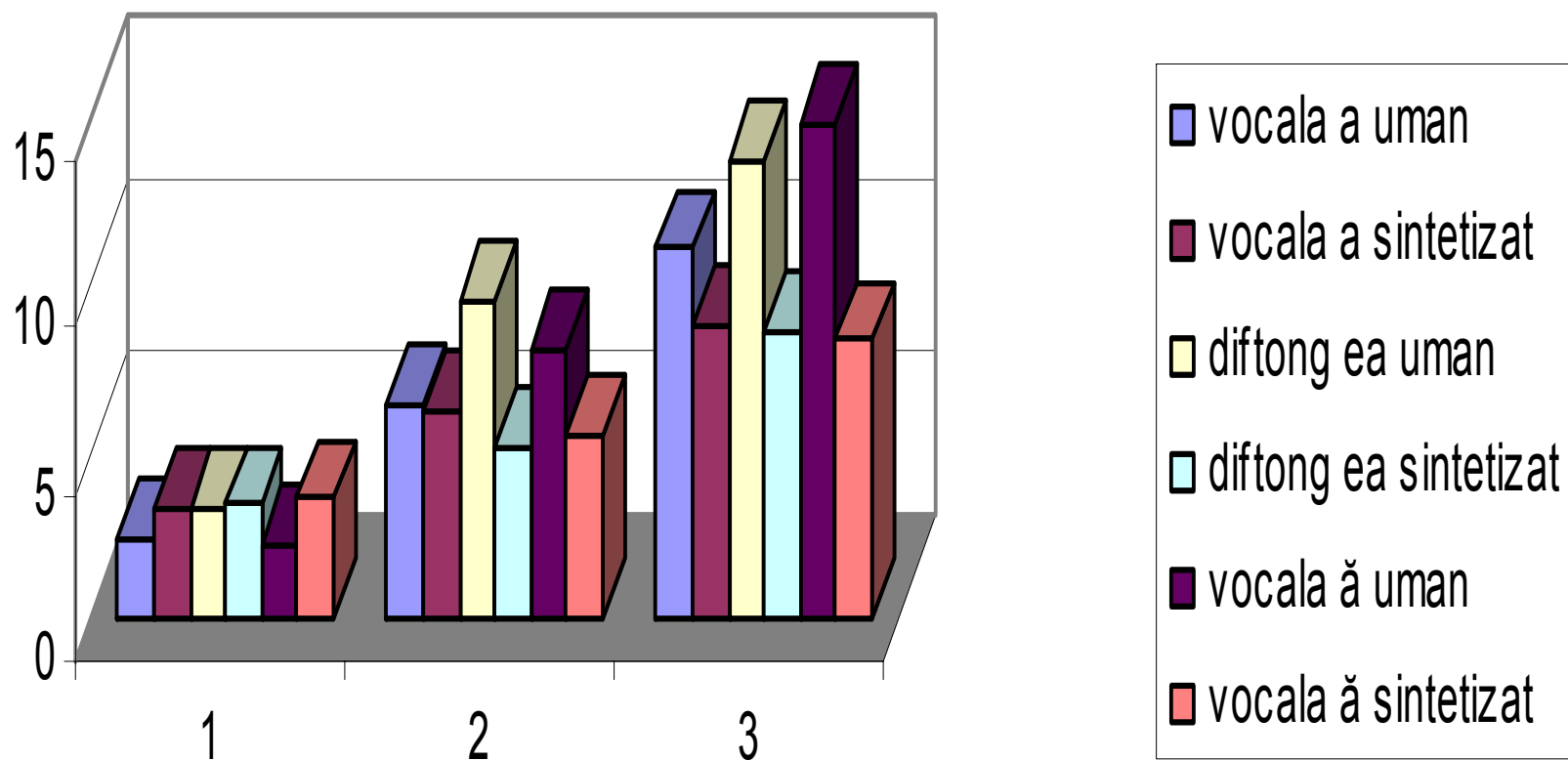
# Rezultatele analizei / “Aseară” (II)

*Aseara* - F0 si formantii F1, F2, F3, valori absolute



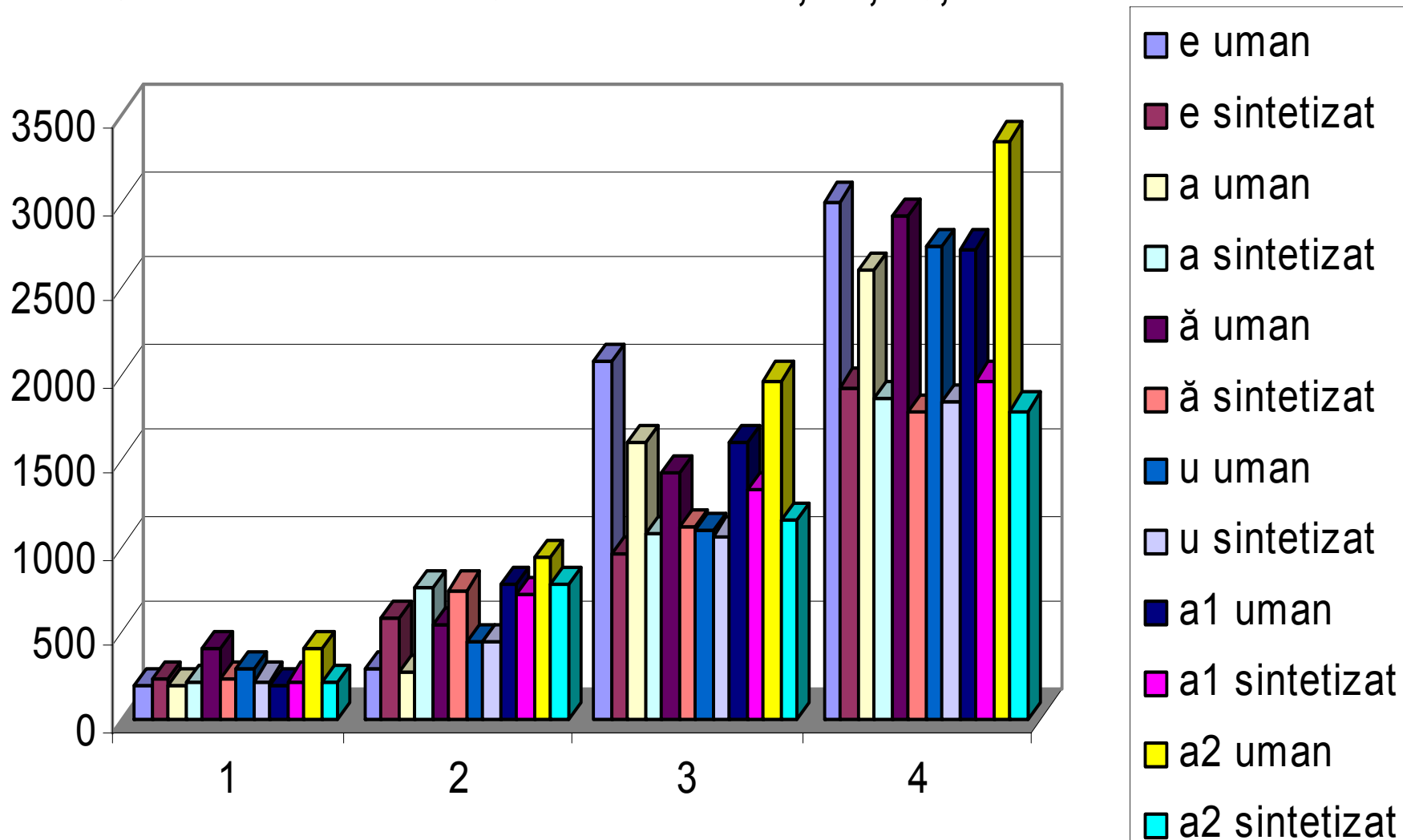
# Rezultatele analizei / “Aseară” (III)

Aseara - raporturi F1/F0, F2/F0 si F3/F0



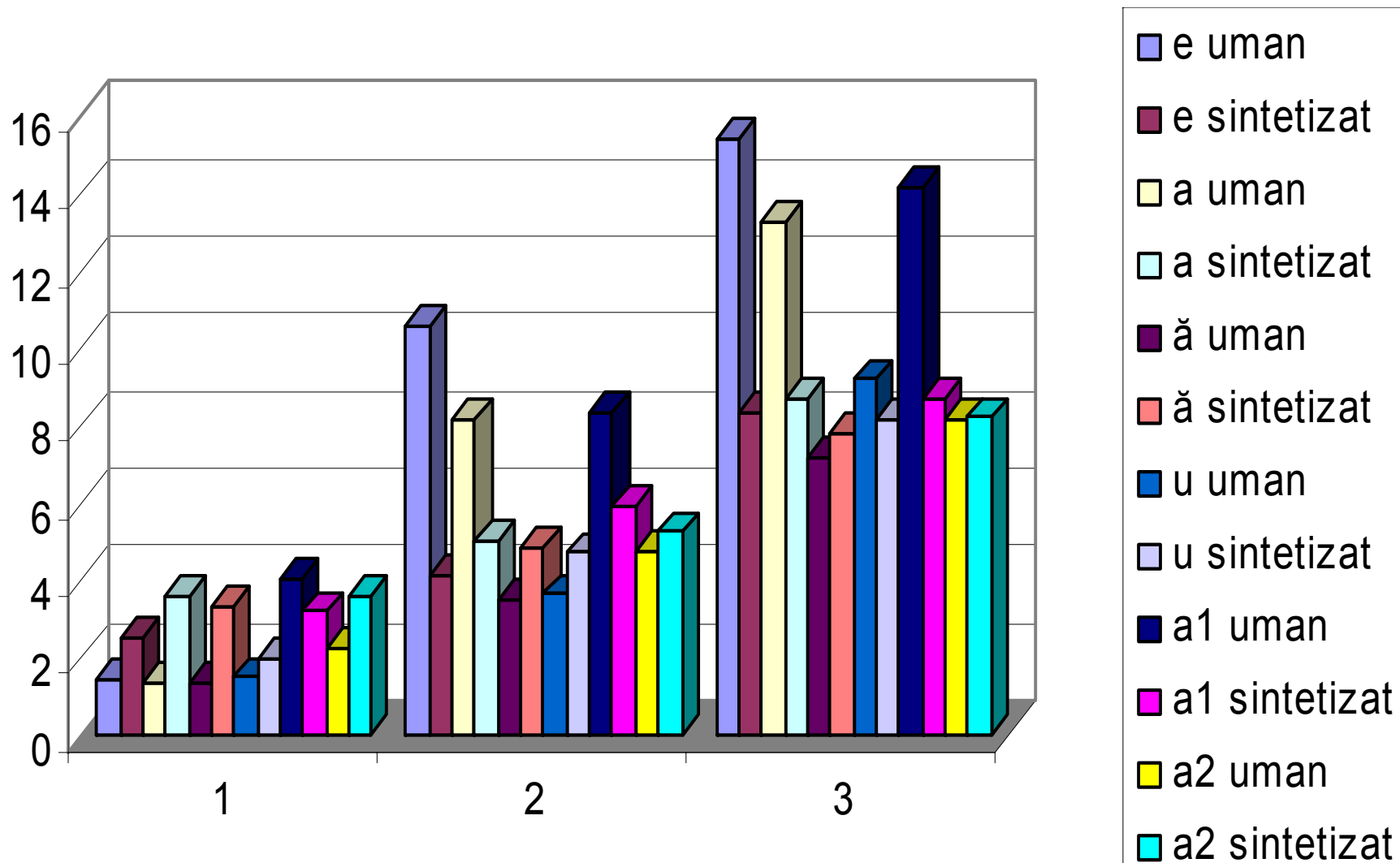
# Rezultatele analizei / “Cine a făcut asta”

*Cine a facut asta* - F0 si formantii F1, F2, F3, valori absolute

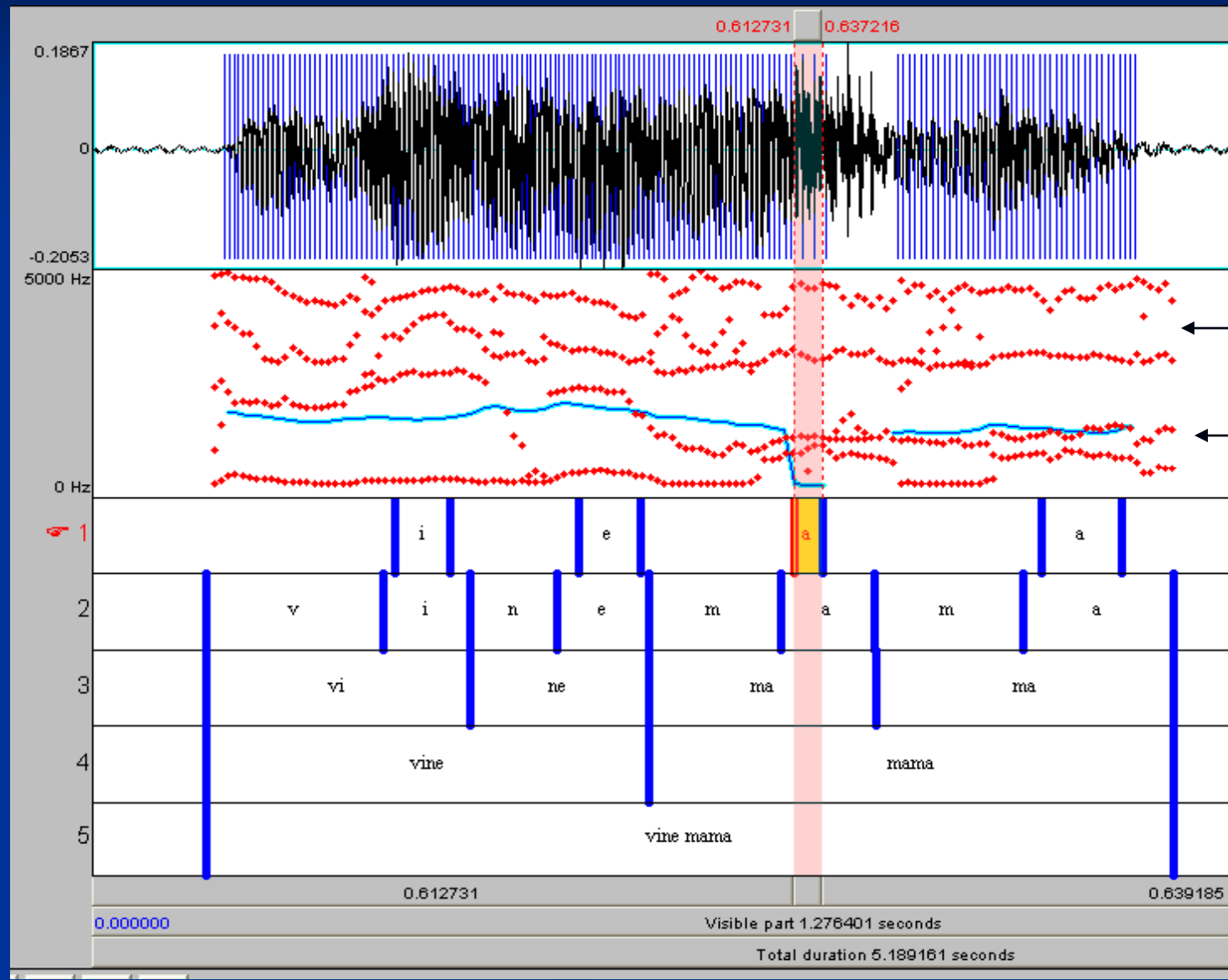


# Rezultatele analizei / “Cine a făcut asta”

*Cine a facut asta?* - raporturi F1/F0, F2/F0 si F3/F0



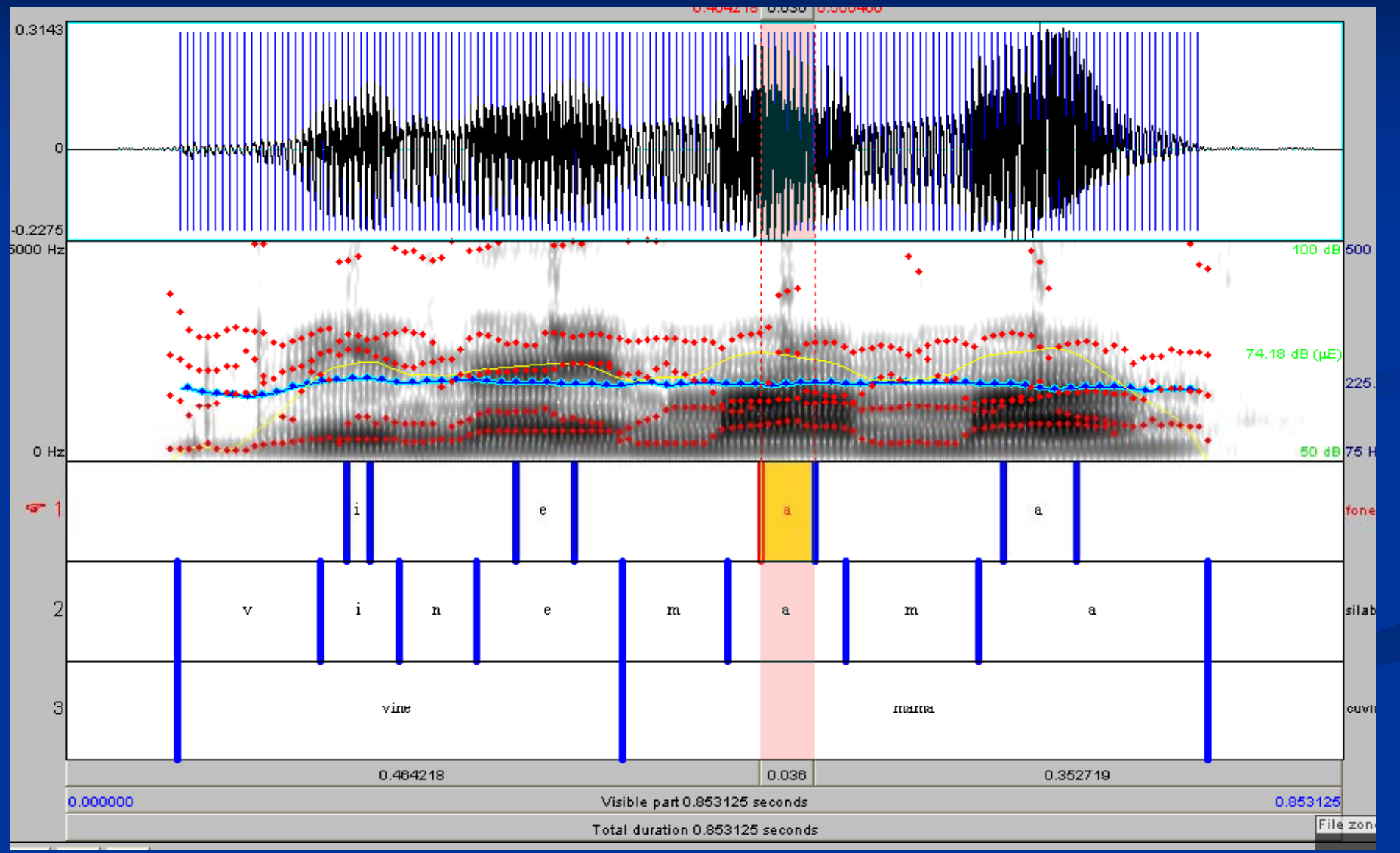
# Exemplu de adnotare manuală folosind utilitarului Praat; voce umană – propoziția “Vine mama”



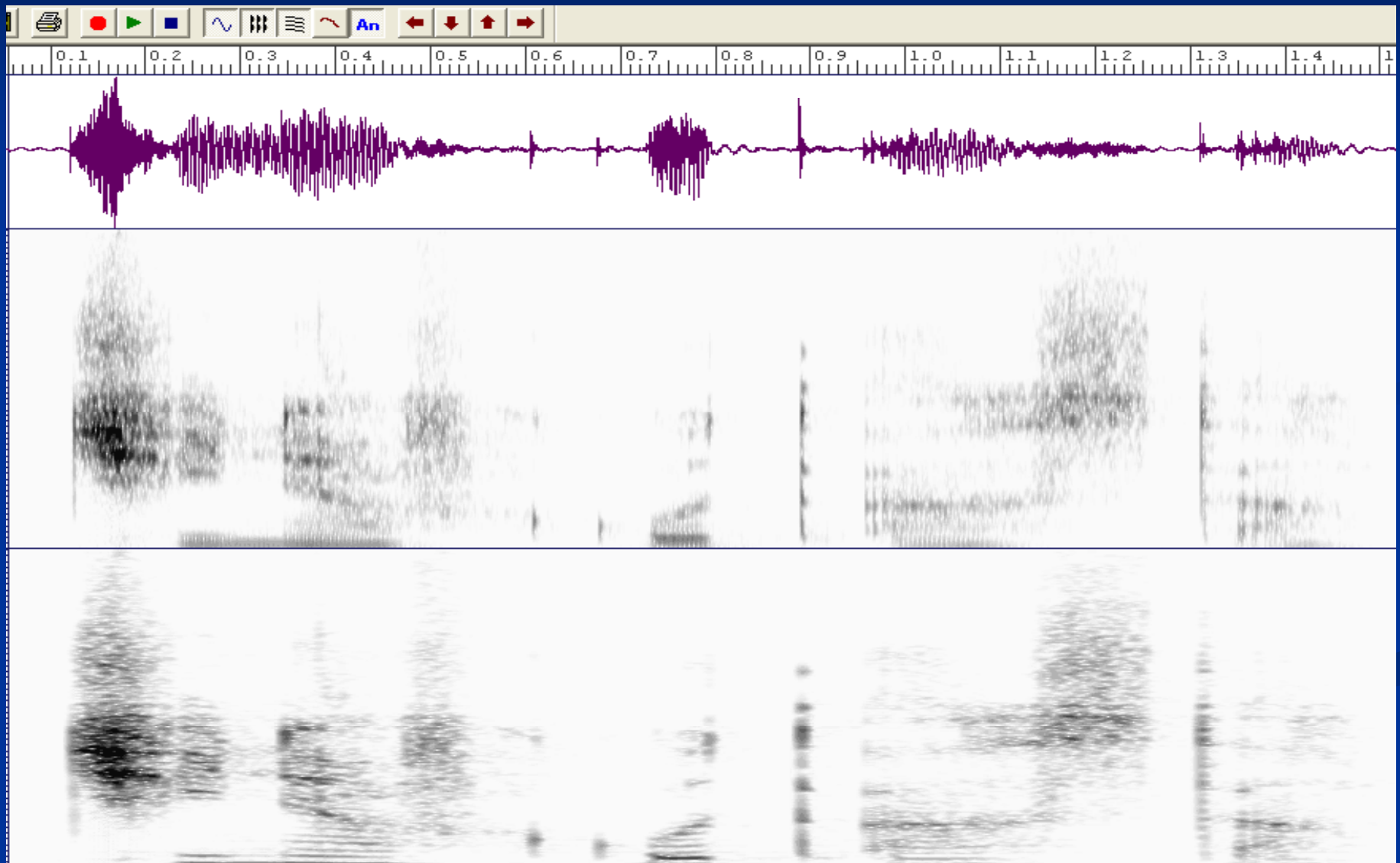
F3, F4 bine demarcati

Variabilitate bogata a traseelor

# Exemplu de adnotare manuală folosind utilitarului Praat; voce sintetizată – propoziția “Vine mama”

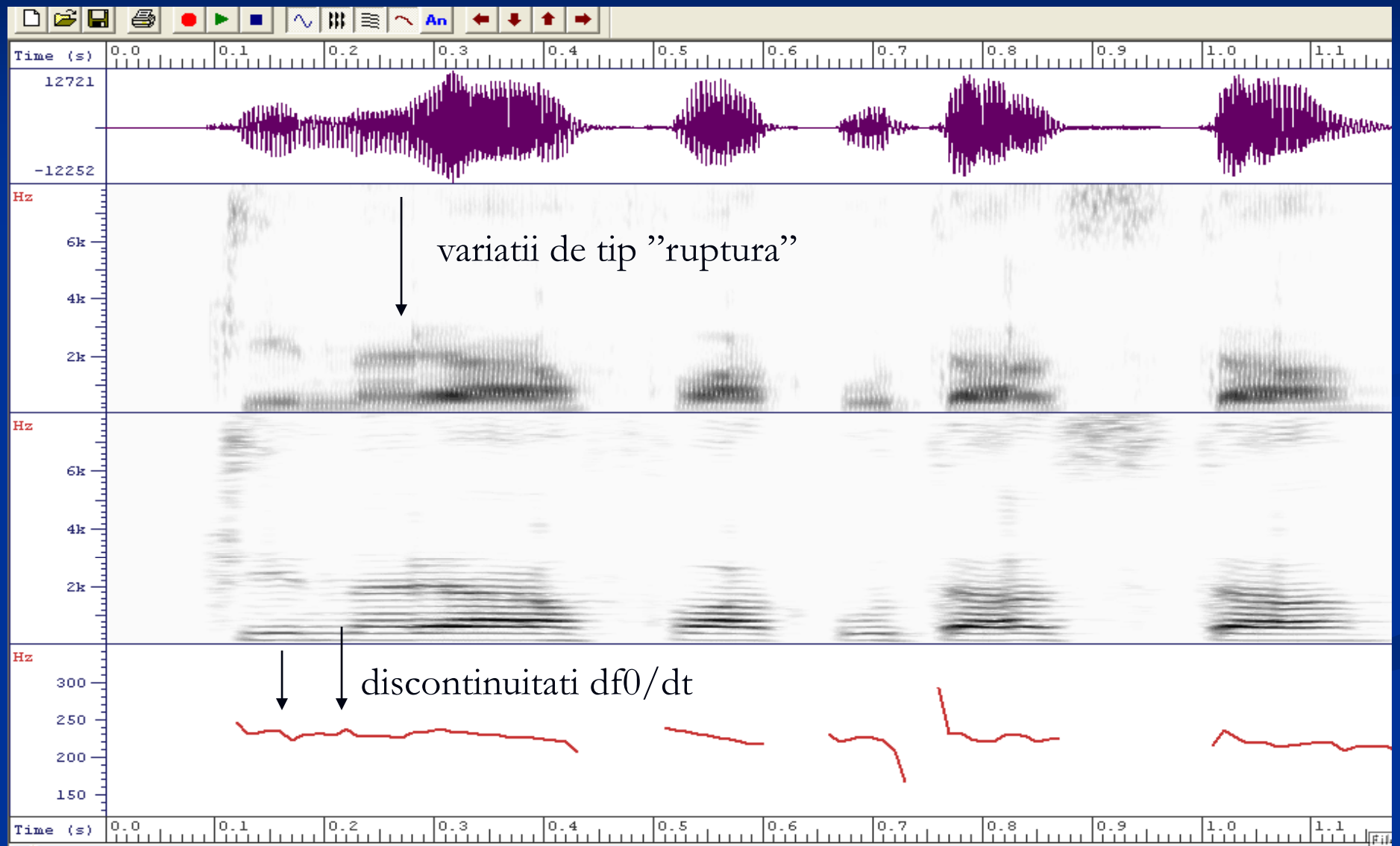


# Traseul frecvenței fundamentale folosind utilitarul Wasp; voce umană - propoziție “Cine a făcut asta”

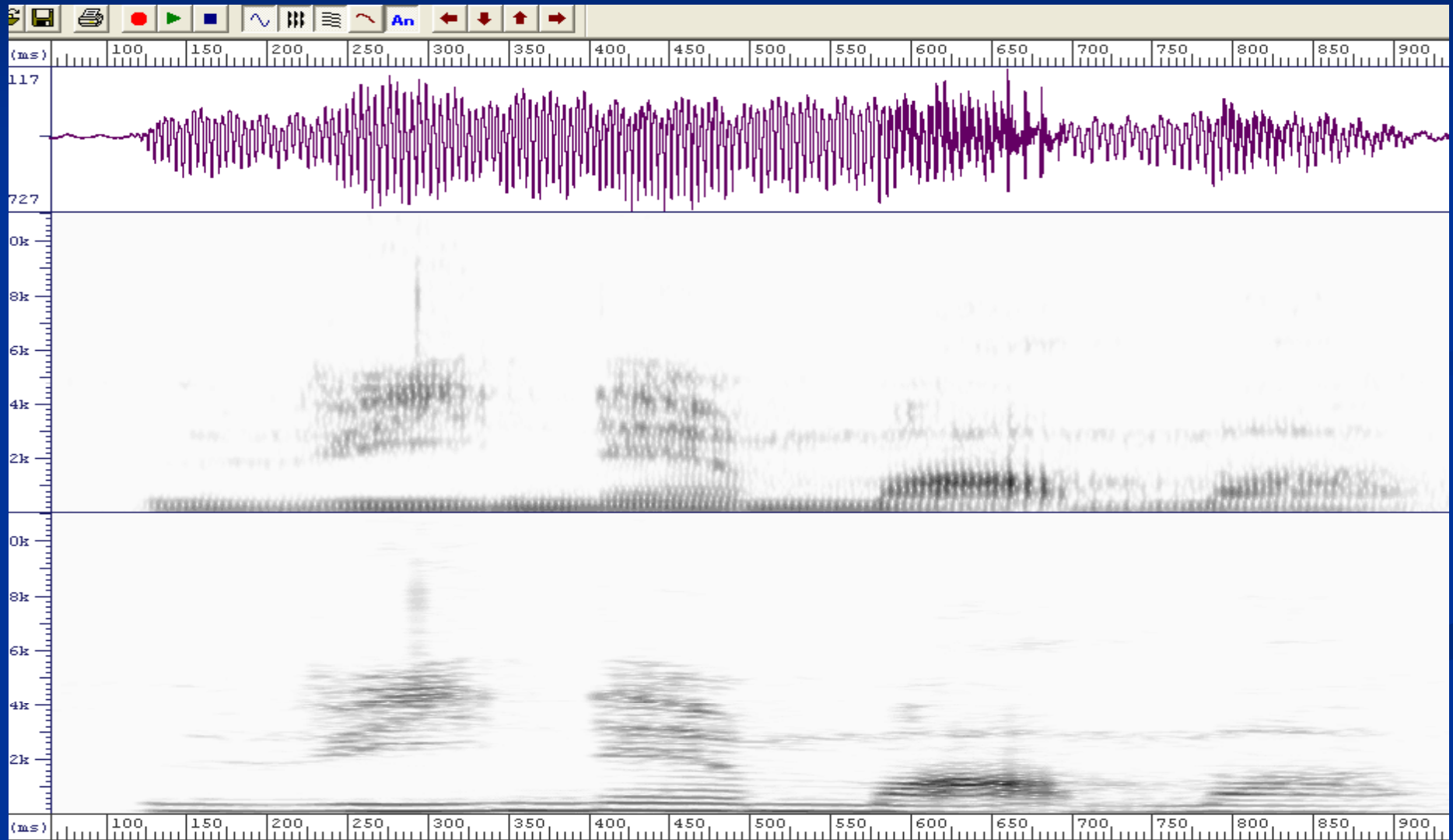




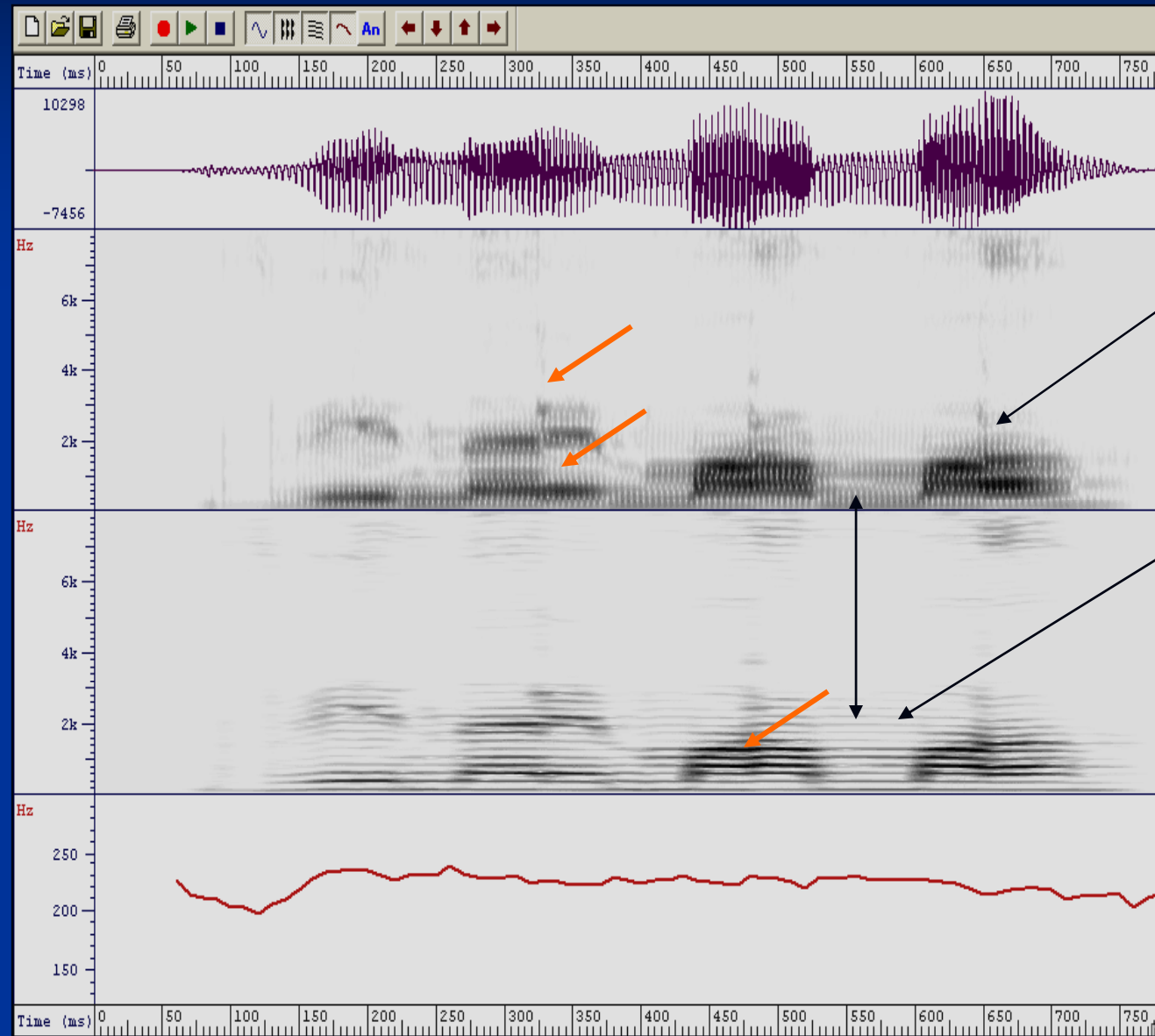
# Traseul frecvenței fundamentale folosind utilitarul Wasp; voce sintetizată - propoziție “Cine a făcut asta”



# Traseul frecvenței fundamentale folosind utilitarul Wasp; voce umană - propoziție “Vine mama”



# Traseul frecvenței fundamentale folosind utilitarul Wasp; voce sintetizată - propoziție “Vine mama”



Benzi  
inguste

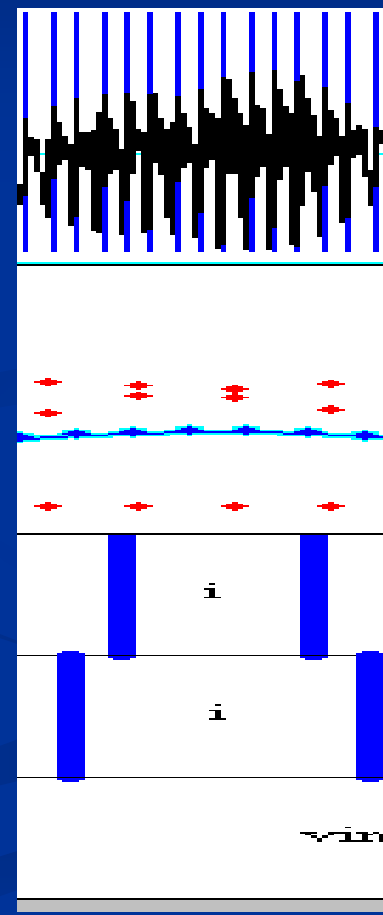
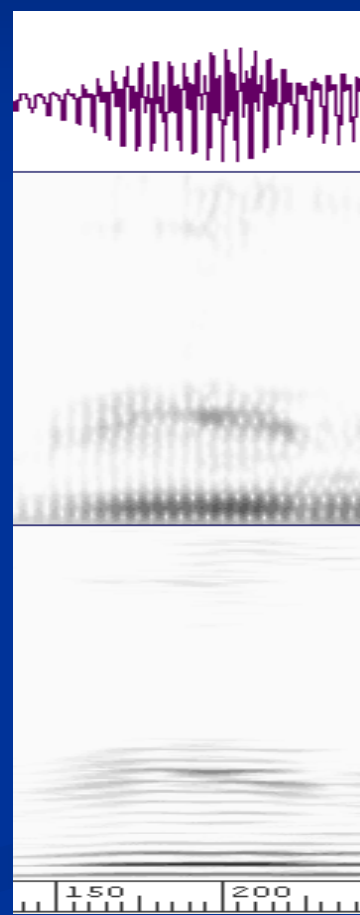
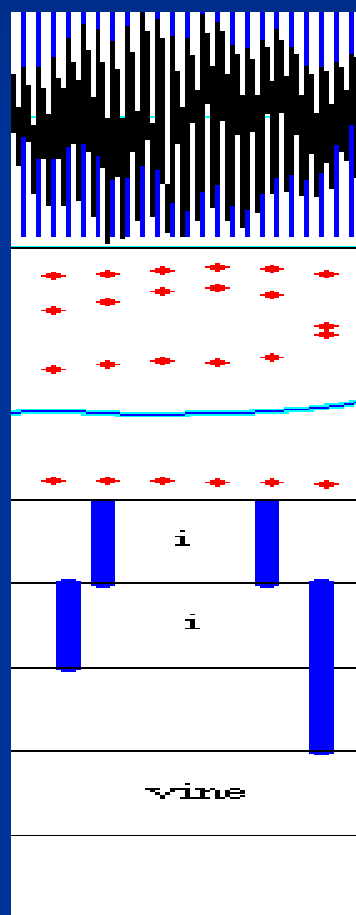
Trasee  
constante,  
fara dinamica

# Comparații între vocea umană și cea sintetică – segment selectat vocala

“i”

voce umană

voce sintetică

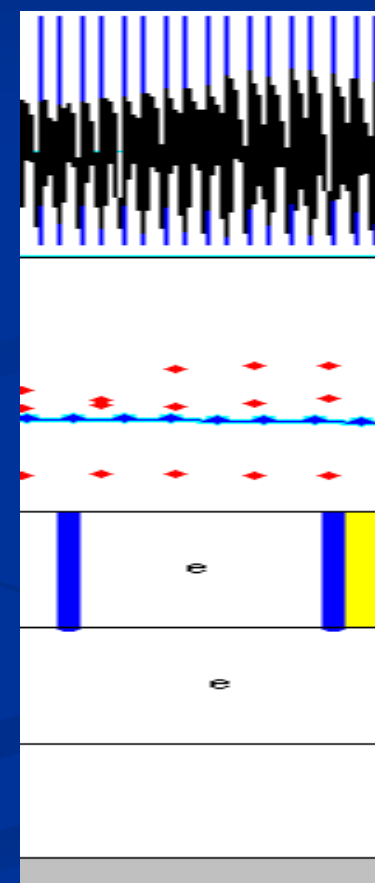
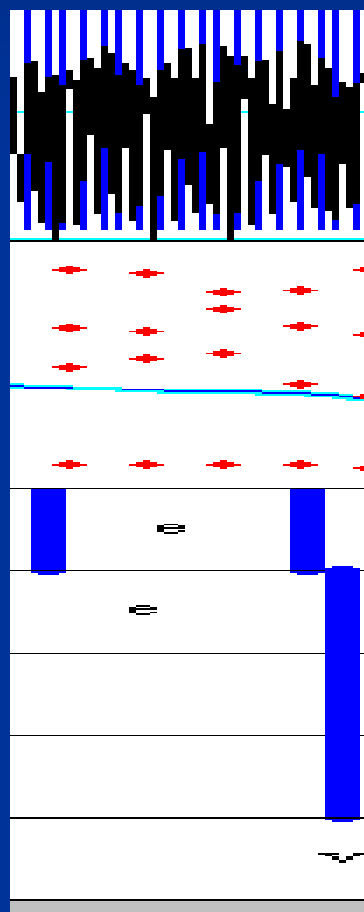


# Comparații între vocea umană și cea sintetică – segment selectat vocala

“e”

voce umană

voce sintetică

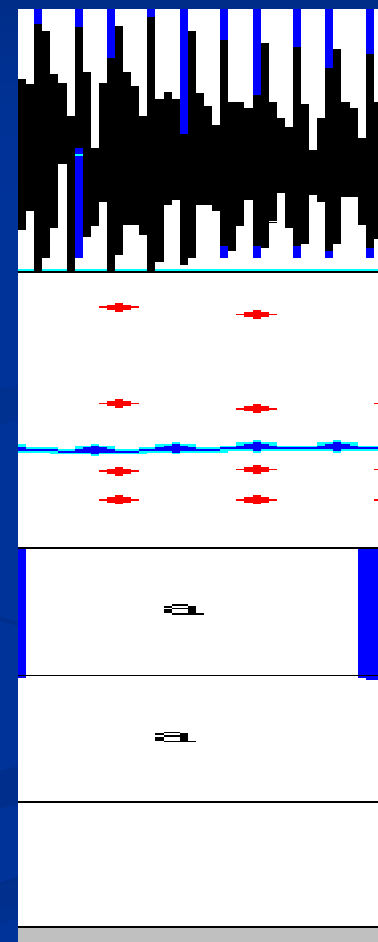
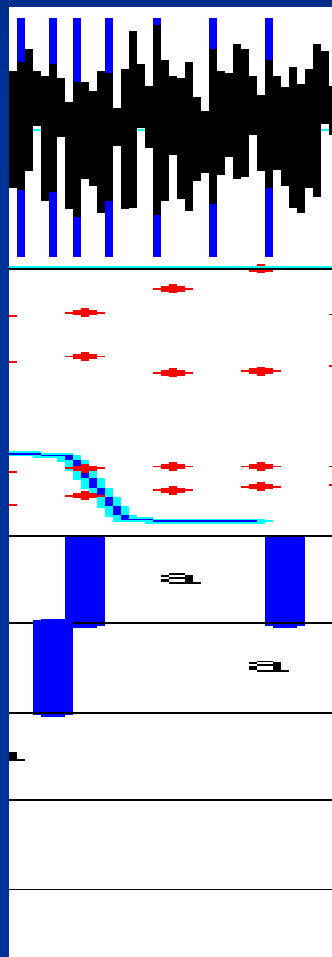
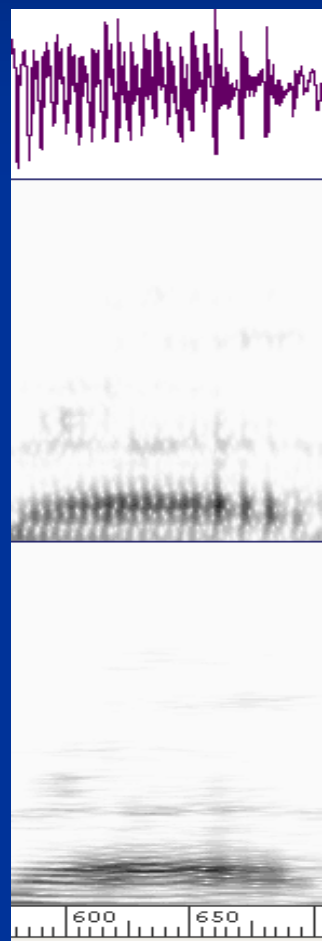


# Comparații între vocea umană și cea sintetică – segment selectat vocala

“a1”

voce umană

voce sintetică

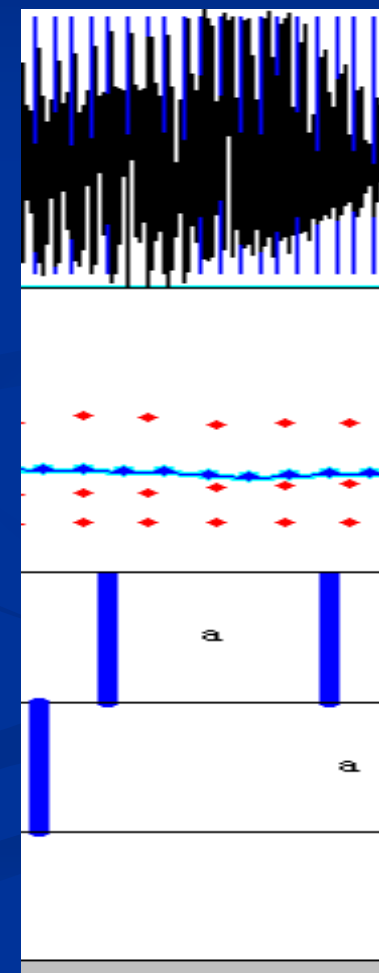
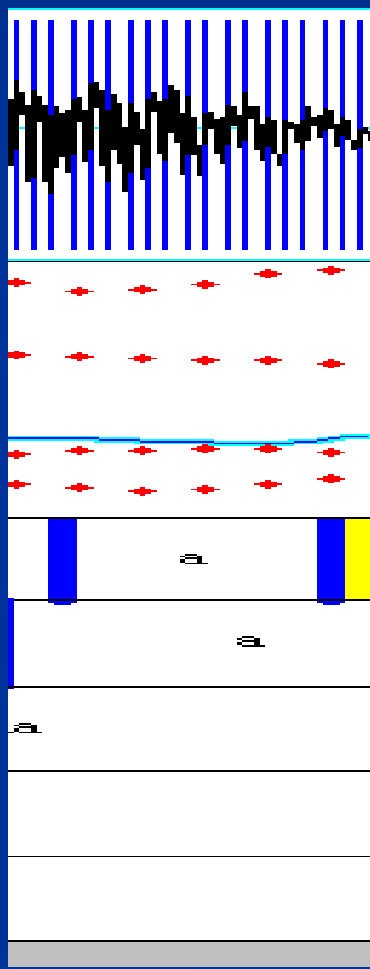
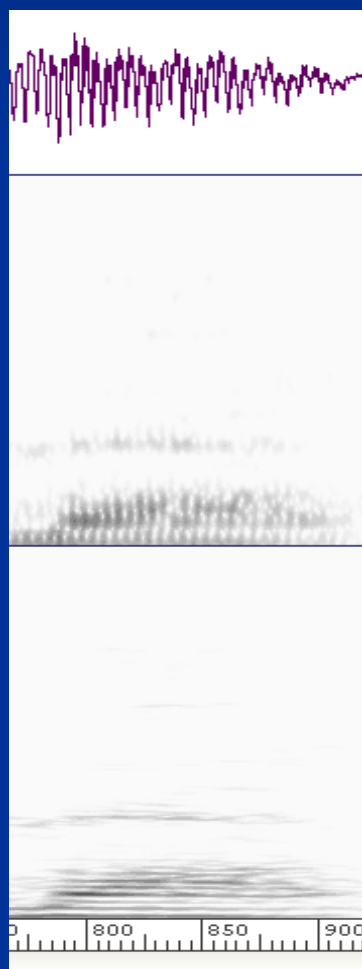


# Comparații între vocea umană și cea sintetică – segment selectat vocala

“a2”

voce umană

voce sintetică



# Concluzii

- am arătat o metodă de verificare cantitativă a “naturaletii” vocii
- o serie de diferențe majore între formanții superiori la vocea sintetică față de cea umană



# Direcții viitoare

- Creșterea naturaleții necesită, foarte probabil, un număr mai mare de fragmente (elemente) în memorie, astfel încât selecția să se facă și ținând cont de traseele pentru formanții superiori.
- Ne propunem în viitor să facem un număr mai mare de comparații pentru limba română – folosind diverse setări pentru F0 pentru sintetizorul respectiv și folosind mai multe voci naturale corespunzător alese.

Vă mulțumesc  
pentru atenție!