

# 음성인식

## 모듈 II

변용혁

장혜원

황수영

박소진

# CONTENTS

## 모듈 II

- 개요  
문제 상황 분석  
접근 방법 & 생각  
개발환경
- 본론  
데이터 분석  
특징 선택 과정  
Decision Tree
- 결과  
테스트 및 오류 수정  
결과  
문제점 분석

## 문제 상황

0 1 2 3 4  
5 6 7 8 9

- 성인 여성의 목소리
- 0.5초 길이 음성
- 특정 크기의 볼륨

## 접근 방법

0 1 2 3 4  
5 6 7 8 9

구별하기 위한 특징

스펙트로그램, 피치, 포먼트, 주파수...

## 개발 환경

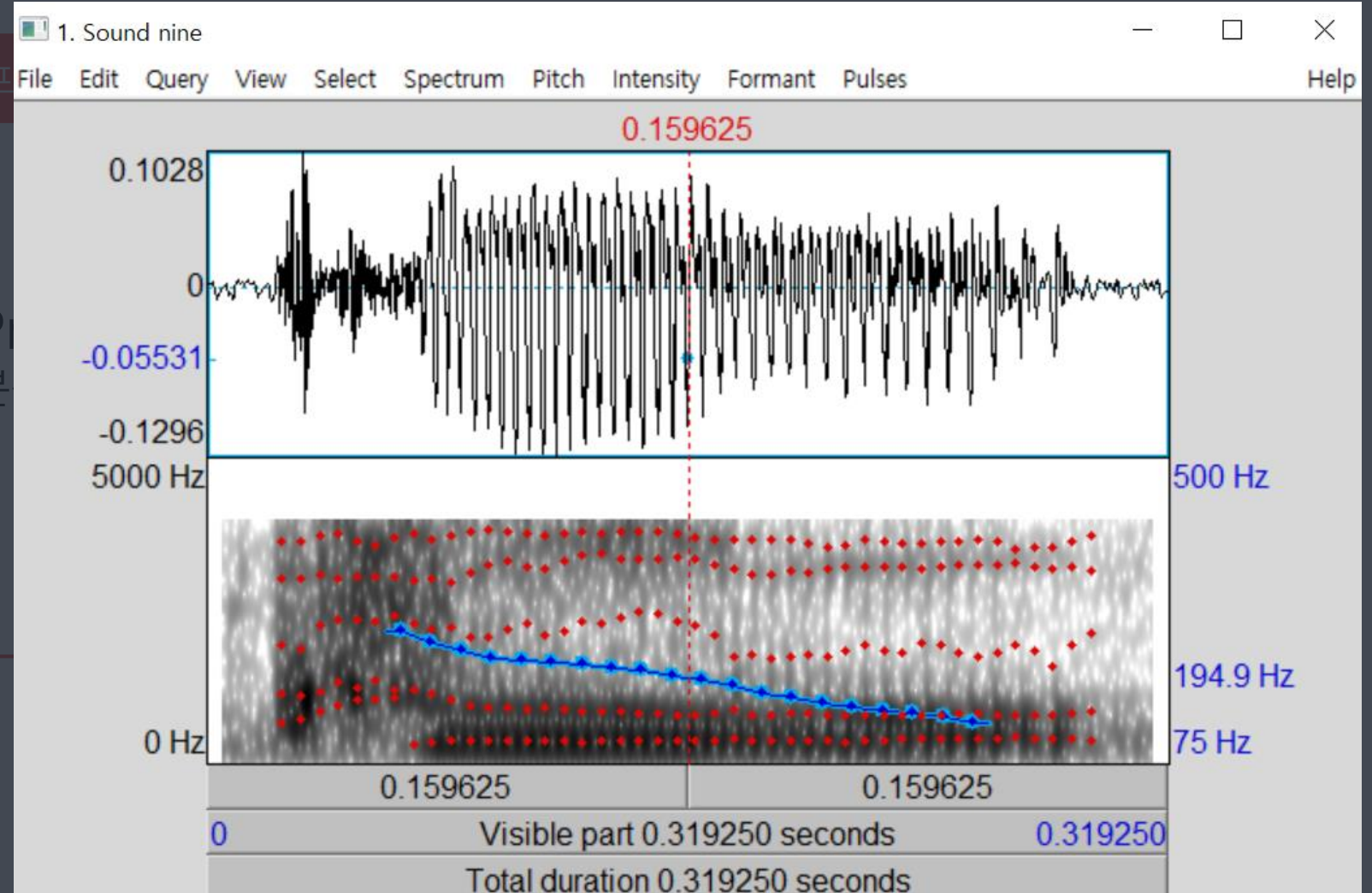
프로그램

Praat  
음성 분석 프로그램

## 개발 환경

## 음성 분석 프로그램

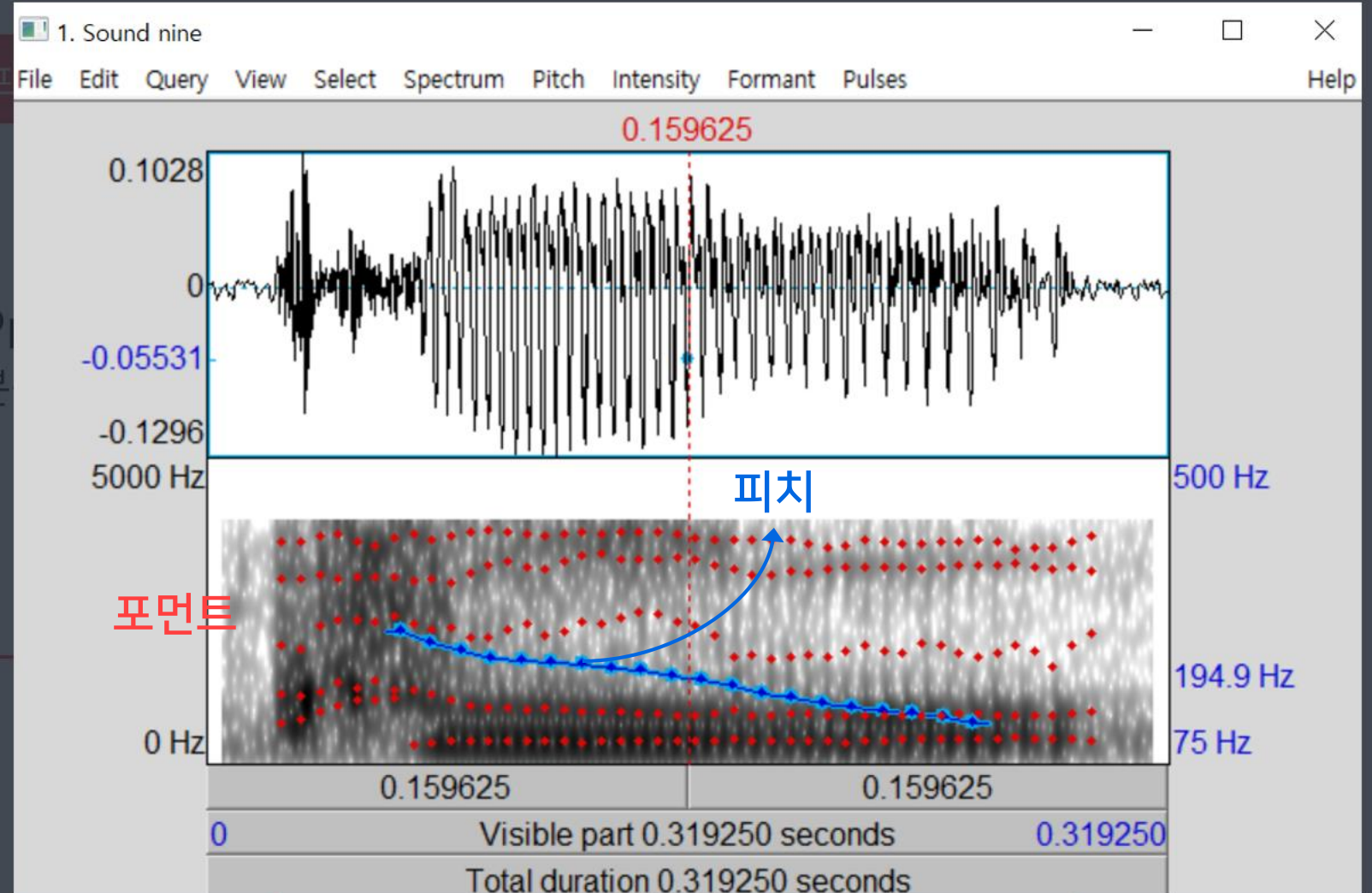
음성 분석



## 개발 환경

### 음성 분석 프로그램

음성 분석



## 개발 환경

### 프로그램

Praat  
음성 분석 프로그램

### 라이브러리

Parselmouth  
프라트 프로그램을 위한  
파이썬 라이브러리

- Maximum formant = 5500Hz
- Number of formants = 4
- window length = 0.025
- time\_step = 0.0625



## 데이터 분석

0 1 2 3 4  
5 6 7 8 9

프라트 분석 프로그램으로  
본 데이터 분석

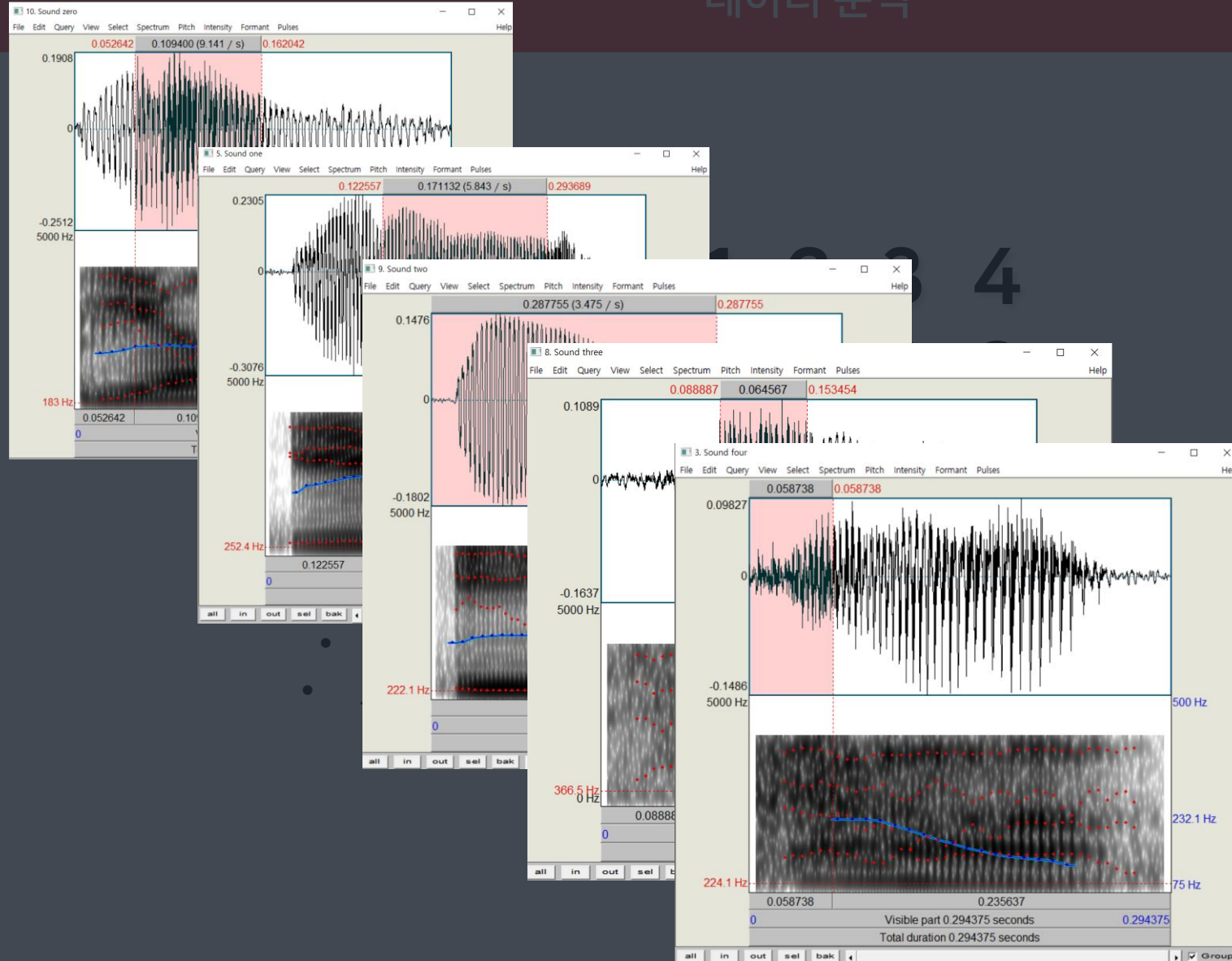
- 숫자 수치
- 그래프 형태

이론 중심으로 본  
데이터 분석

- 입벌림 정도에 대한  
포먼트 값
- 고모음과 저모음  
포먼트 차이

# 데이터 분석

## 그래프 분석



# 데이터 분석

파일										홈	삽입	페이지 레이아웃	수식	데이터	검토	보기	도움말	Power Pivot	팀	어떤 작업을 원하시나요?																
T23										2448.589034																										
1	0.02781	425	746	2179	3357	1.8			0.02769	314	2063	2520	3279	6.6	-2.7	0.0275	684	1927	2421	3245			0.027312	674	1446	2267	2989			0.02525	637	1132	1924	2992		
2	0.03406	412	739	2097	3393	1.8			0.03394	276	1979	2585	3320	7.2	-3.3	0.03375	754	1781	2374	3276			0.033562	666	1425	2203	3038			0.0315	689	1095	1859	2997		
3	0.04031	416	769	1821	3315	1.8			0.04019	259	1779	2576	3345	6.9	-3.0	0.04	778	1684	2403	3230			0.039812	705	1437	2206	3040			0.03775	822	1161	2240	3063		
4	0.04556	412	768	1907	3195	1.9			0.04544	245	1664	2734	3366	6.9	-3.0	0.04525	688	1693	2440	3190			0.044062	819	1477	2281	2988			0.044	937	1312	2336	3006		
5	0.05281	412	779	1895	3190	1.9			0.05269	236	1640	2786	3475	6.9	-3.0	0.0525	684	1734	2453	3210			0.052312	960	1616	2270	2913			0.05025	1019	1199	2343	3040		
6	0.05906	416	783	1793	3247	1.9			0.05894	239	1637	2787	3444	6.8	-2.9	0.05875	507	1740	2443	3221		662	0.058562	972	1593	2283	2911			0.0565	1015	1329	2302	3025		
7	0.06531	419	762	1743	3180	1.8			0.06519	238	1631	2734	3462	6.8	-2.9	0.065	376	1823	2462	3173			0.064812	988	1393	2245	3000			0.06275	1042	1172	2400	3029	880	
8	0.07156	425	741	1718	3041	1.7			0.07144	240	1476	2737	3471	6.2	-2.3	0.07125	373	1446	2586	3115			0.071062	956	1332	2317	3012			0.069	292	1180	2268	2975		
9	0.07781	424	726	1610	3034	1.7			0.07769	245	1392	2740	3324	5.7	-1.8	0.0775	404	1121	2610	3039			0.077312	947	1292	2342	3020			0.07525	317	1121	2180	2995		
10	0.08406	417	716	1532	2984	1.7			0.08394	249	1351	2742	3283	5.4	-1.5	0.08375	426	1024	2586	2937			0.083562	950	1344	2344	2983			0.0815	338	1037	2194	2946		
11	0.09031	411	718	1550	3004	1.7			0.09019	257	1262	2720	3270	4.9	-1.0	0.09	427	1045	2508	2966			0.089912	995	1203	2296	3045			0.08775	339	919	2048	3096		
12	0.09656	412	717	1517	2971	1.7			0.09644	272	1180	2650	3155	4.3	-0.4	0.09625	436	1129	2557	2982			0.096062	756	1073	2274	3048			0.094	339	898	2055	3208		
13	0.10281	408	720	1480	2902	1.8			0.10269	297	1124	2591	3076	3.8	0.1	0.1025	432	1182	2607	2982			0.102313	1078	1235	2972	3528	881		0.10025	337	880	2174	3271		
14	0.10906	400	722	2005	2910	1.8			0.10894	340	1144	2566	3089	3.4	0.5	0.10875	411	1176	2620	3013			0.108562	586	1062	2275	2934			0.1065	342	872	2260	3205		
15	0.11531	439	679	2082	2866	1.5			0.11519	382	1182	2502	3061	3.1	0.8	0.115	426	1078	2574	3048			0.114813	761	990	2367	2944			0.11275	343	883	2083	3170		
16	0.12156	522	639	2223	2760	1.2			0.12144	416	1221	2394	3024	2.9	1.0	0.12125	443	1059	2596	3103			0.121062	705	1002	2358	3059			0.119	339	865	2148	3282		
17	0.12781	549	560	2144	2735	1.0			0.12769	432	1268	2352	2934	2.9	1.0	0.1275	431	1073	2654	3160			0.127313	562	1022	2465	3112			0.12525	333	845	2305	3388		
18	0.13406	382	545	2039	2724	1.4			0.13394	387	1349	2324	2902	3.1	0.8	0.13375	445	1002	2567	3186			0.133563	488	1030	2412	3091			0.1315	338	834	2271	3426		
19	0.14031	542	684	1946	2711	1.3			0.14019	443	1439	2213	2968	3.2	0.7	0.14	456	1310	2537	3153			0.139813	436	1021	2412	2990			0.13775	348	833	2372	3326		
20	0.14656	426	574	1901	2670	1.3			0.14644	445	1486	1986	3007	3.3	0.6	0.14625	466	1276	2531	3115			0.146062	490	998	2423	3060			0.144	346	811	2441	3329		
21	0.15281	552	600	1889	2647	1.1			0.15269	444	1548	1724	2988	3.5	0.4	0.1525	430	1234	2455	3222			0.152312	564	989	2445	3139			0.15025	344	794	2432	3330		
22	0.15906	526	616	1877	2704	1.2			0.15894	440	1549	1566	3002	3.5	0.4	0.15875	443	1244	2447	3245			0.158563	529	965	2416	3120			0.1565	341	778	2299	3365		
23	0.16531	513	622	1893	2697	1.2			0.16519	424	1425	1553	3051	3.4	0.5	0.165	453	1279	2443	3219			0.164913	406	934	2444	3121			0.16275	346	765	2230	3327		
24	0.17156	586	602	1903	2673	1.0			0.17144	404	1252	1592	3043	3.1	0.8	0.17125	430	1283	2359	3180			0.171063	365	914	2494	3169			0.169	364	787	2065	3230		
25	0.17781	571	697	1673	2654	1.2			0.17769	395	1210	1560	3068	3.1	0.8	0.1775	437	1202	2479	3203			0.177312	209	885	2470	3177			0.17525	382	859	1731	3154		
26	0.18406	534	780	1821	2654	1.5			0.18394	391	1198	1538	3080	3.1	0.8	0.18375	403	1275	2434	3180			0.183562	862	2450	3118	3619			0.1815	360	781	1739	3058		
27	0.19031	505	884	1812	2678	1.8			0.19019	382	1160	1608	3061	3.0	0.9	0.19	399	1369	2436	3191			0.189813	858	2467	3109	3614			0.18775	364	768	1685	3051		
28	0.19656	493	824	1821	2691	1.7			0.19644	397	1146	1665	3092	2.9	1.0	0.19625	407	1432	2429	3183			0.196063	274	827	2472	3103			0.194	360	798	1711	3124		
29	0.20281	485	773	1923	2691	1.6			0.20269	397	1110	1520	3004	2.8	1.1	0.2025	406	1485	2413	3166			0.202313	365	862	2443	3174			0.20025	347	798	1771	3110		
30	0.20906	463	787	2091	2755	1.7			0.20894	394	1077	1632	3122	2.7	1.2	0.20875	403	1430	2405	3167			0.208563	433	847	2475	3108			0.2065	338	790	1731	3150		
31	0.21531	430	779	2048	2970	1.8			0.21519	394	1081	1606	3138	2.7	1.2	0.215	402	1423	2405	3155			0.214812	471	886	2517	3046			0.21275	330	772	1824	3175		
32	0.22156	419	767	2297	3252	1.8			0.22144	380	1063	1599	3145	2.8	1.1	0.22125	402	1473	2399	3151			0.221063	486	948	2510	3135			0.219	346	766	1907	3178		
33	0.22781	418	752	2429	3348	1.8			0.22769	377	1062	1592	3208	2.8	1.1	0.2275	405	1474	2371	3109			0.227313	491	1068	2524	3167			0.22525	362	757	1822	3190		
34	0.23406	419	737	2562	3316	1.8			0.23394	392	1030	1525	3230	2.6	1.3	0.23375	423	1498	2371	3166	421		0.233563	483	1140	2543	3112			0.2315	376	755	1779	3187		
35	0.24031	417	721	2605	3357	1.7			0.24019	396	958	1612	3159	2.7	1.2	0.24	443	1544	2386	3225			0.239813	488	1202	2541	3151			0.23775	389	756	1936	3166		
36	0.24656	406	709	2563	3392	1.7			0.24644	239	889	1579	3192	3.7	0.2	0.24625	452	1618	2378	3241			0.246062	491	1228	2529	3150			0.244	390	767	1906	3184		
37	0.25281	415	719	2544	3384	1.7			0.25269	396	1665	1609	3392	4.2	-0.3	0.2525	456	1609	2358	3258			0.252313	490	1291	2538	3195			0.25025	382					

## 사용한 특징

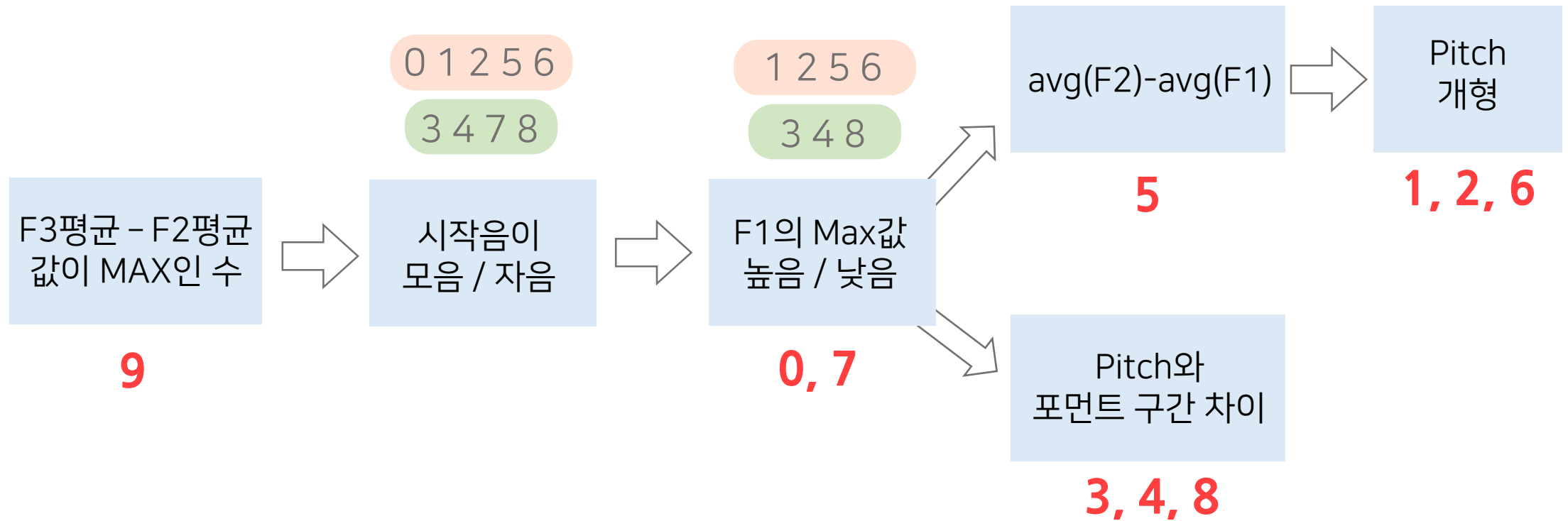
### ▶ 포먼트 값 formant value

- F2평균 - F1평균
- F3평균 - F2평균
- F1의 Max

### ▶ 피치 값 pitch value

- **포먼트 구간** 정의  
모음 시작 그룹, 자음시작 그룹 구분
- 피치 그래프 개형

## 흐름도



0 1 2 3 4 5 6 7 8 9

0 1 2 3 4 5 6 7 8

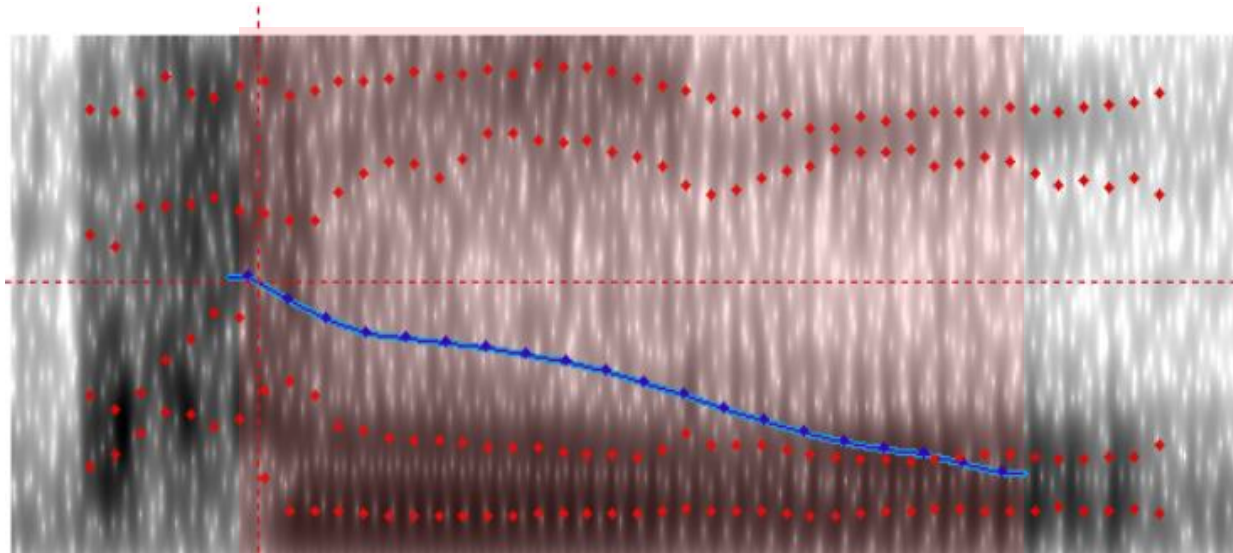
9



Group 0

F3 - F2 의 평균이 가장 큰 값

→ 9 구분



9의 스펙트로그램

## 그룹화

F3평균 - F2평균값이 MAX인 수 → 9 Group 0

모음/자음	0
	모음
Max(F1) 700Hz 기준	높

Group 1

7
자음
낮

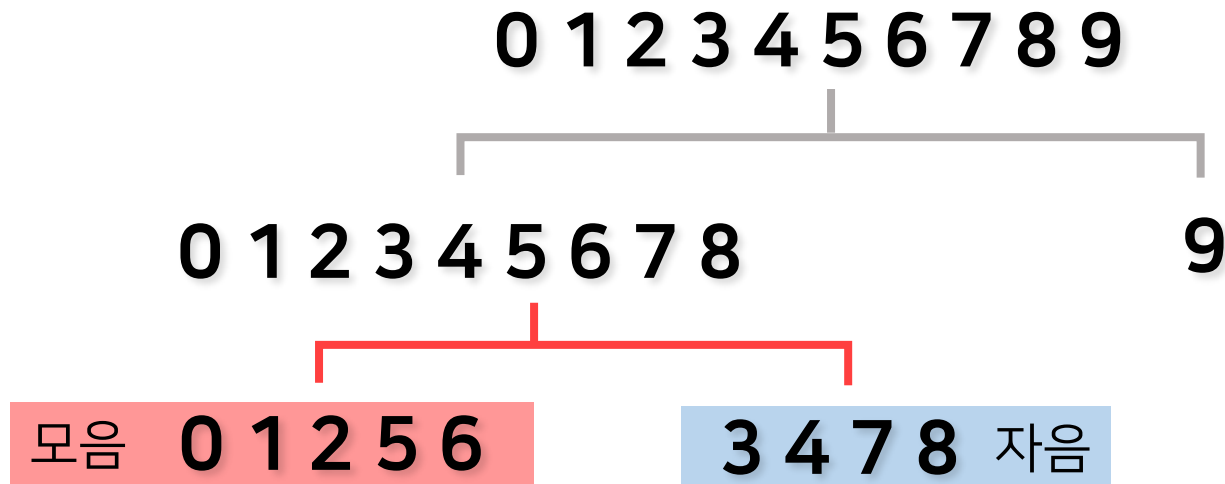
Group 2

1	2	5	6
모음	모음	모음	모음
낮	낮	낮	낮

Group 3

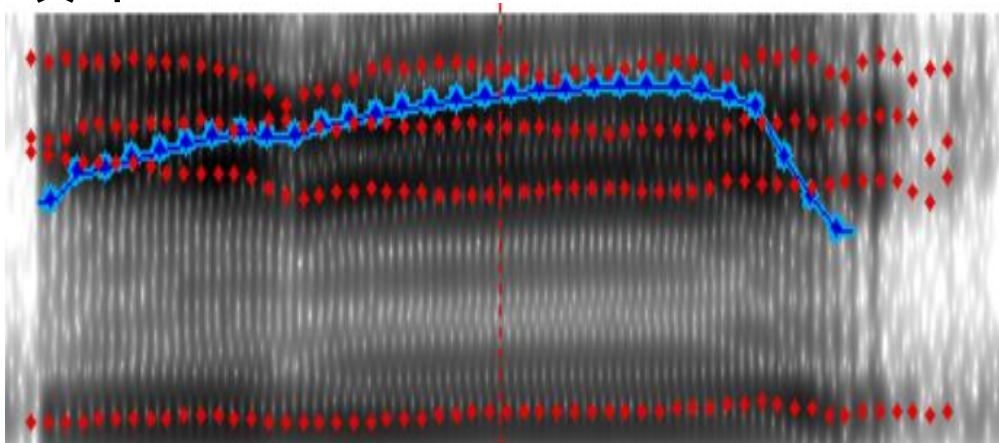
3	4	8
자음	자음	자음
높	높	높

Group 4



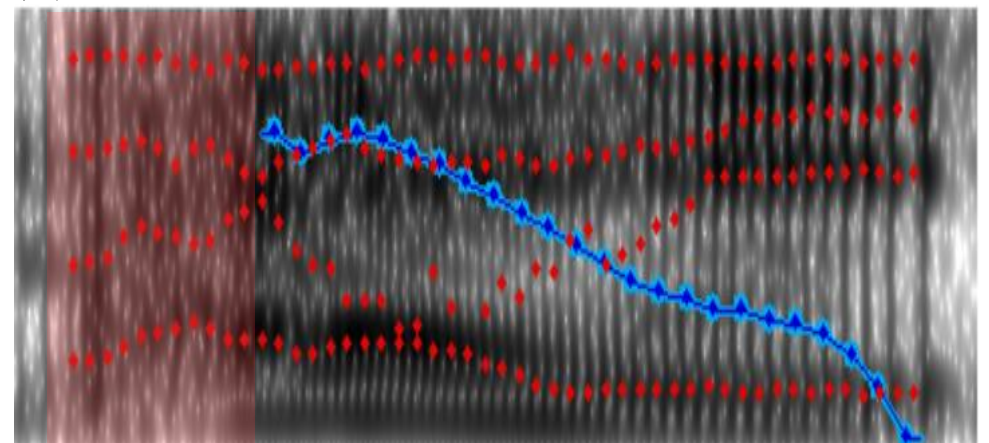
- 모음 시작 / 자음 시작  
Praat 분석프로그램 상에서 Pitch가 정의되는 구간과 포먼트가 정의되는 구간의 시작점의 차이를 이용함

숫자 0



Pitch가 정의되는 구간과 포먼트가 정의되는 구간이 유사함

숫자 8



Pitch가 정의되는 구간과 포먼트가 정의되는 구간의 차이가 큼



## ○ 자음 시작 그룹과 모음 시작 그룹을 구분할 때 시행착오

### 문제

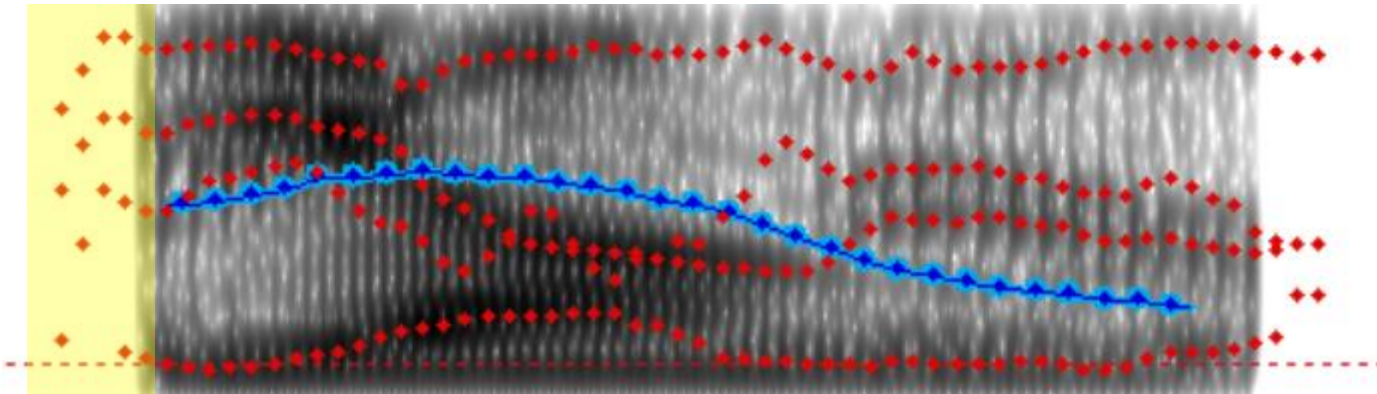
테스트할 때 위치를 옮겼던 음성데이터에서 기존의 음성에서는 나타나지 않았던 포먼트값들이 발생

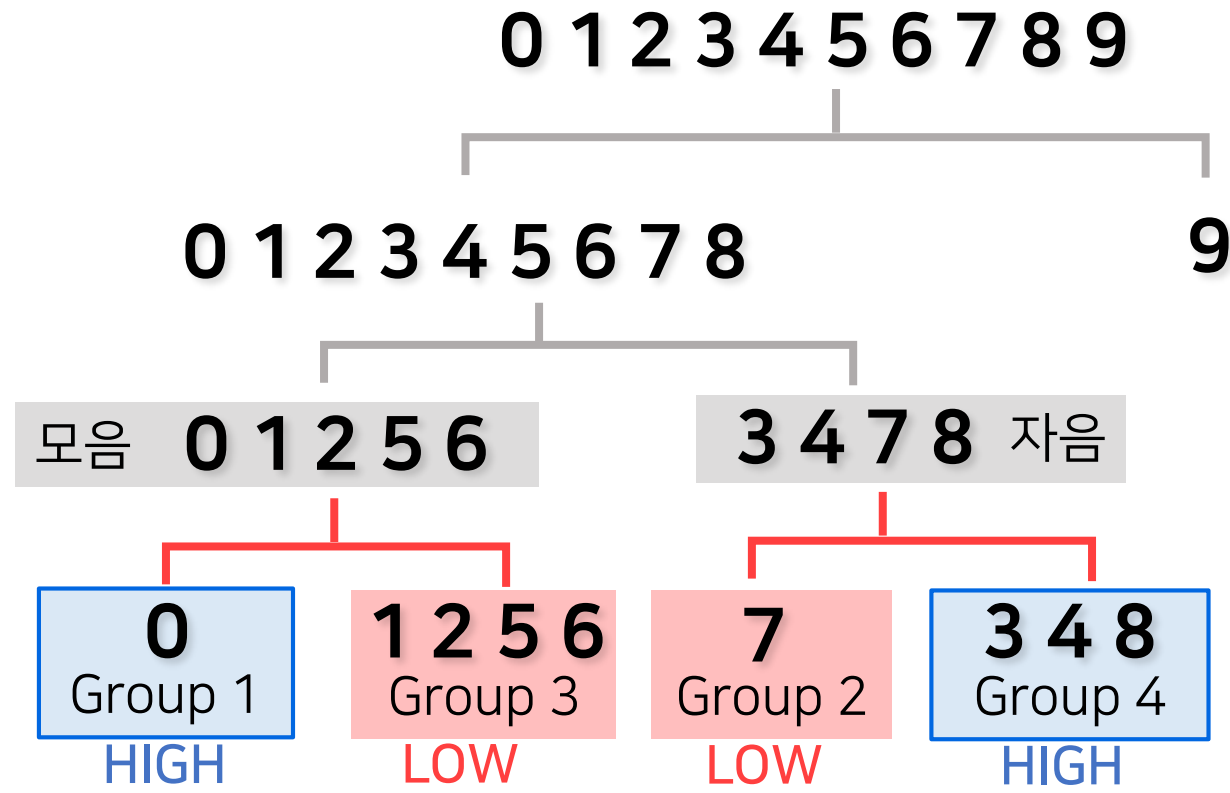
→ 자음과 모음 특징을 초반에 사용했기 때문에 이 후 대부분의 결과에 큰 영향

### 해결책

praat분석 프로그램의 결과에 따라 테스트데이터를 분석해본 결과 모든 변조된 데이터의 앞,뒤 부분에 10개정도의 데이터가 의도적으로 잡혀있음을 확인함

→ 필요한 데이터만 필터링 해 사용





○ F1의 MAX값 임계값 : 700 기준  
높은 그룹과 낮은  
그룹으로 구분  
→ 0, 7 구분 가능

단위 : Hz

Num	Max(F1)
0	855
1	524
2	284
3	946
4	969
5	613
6	500
7	555
8	1128

오류값 제거

## ○ F1의 MAX값을 구할 때 시행착오

### 문제

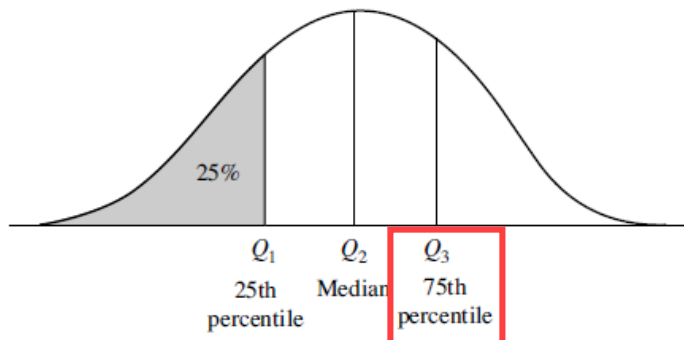
정확치 않은 Pitch 범위로 인해 F1의 max값에서 이상치 존재

→ 그룹1(숫자 0), 그룹2(숫자 7)을 구할 때 잘못된 결과 발생

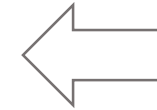
### 해결책

4분위수를 이용해 값의 75퍼센트만 고려

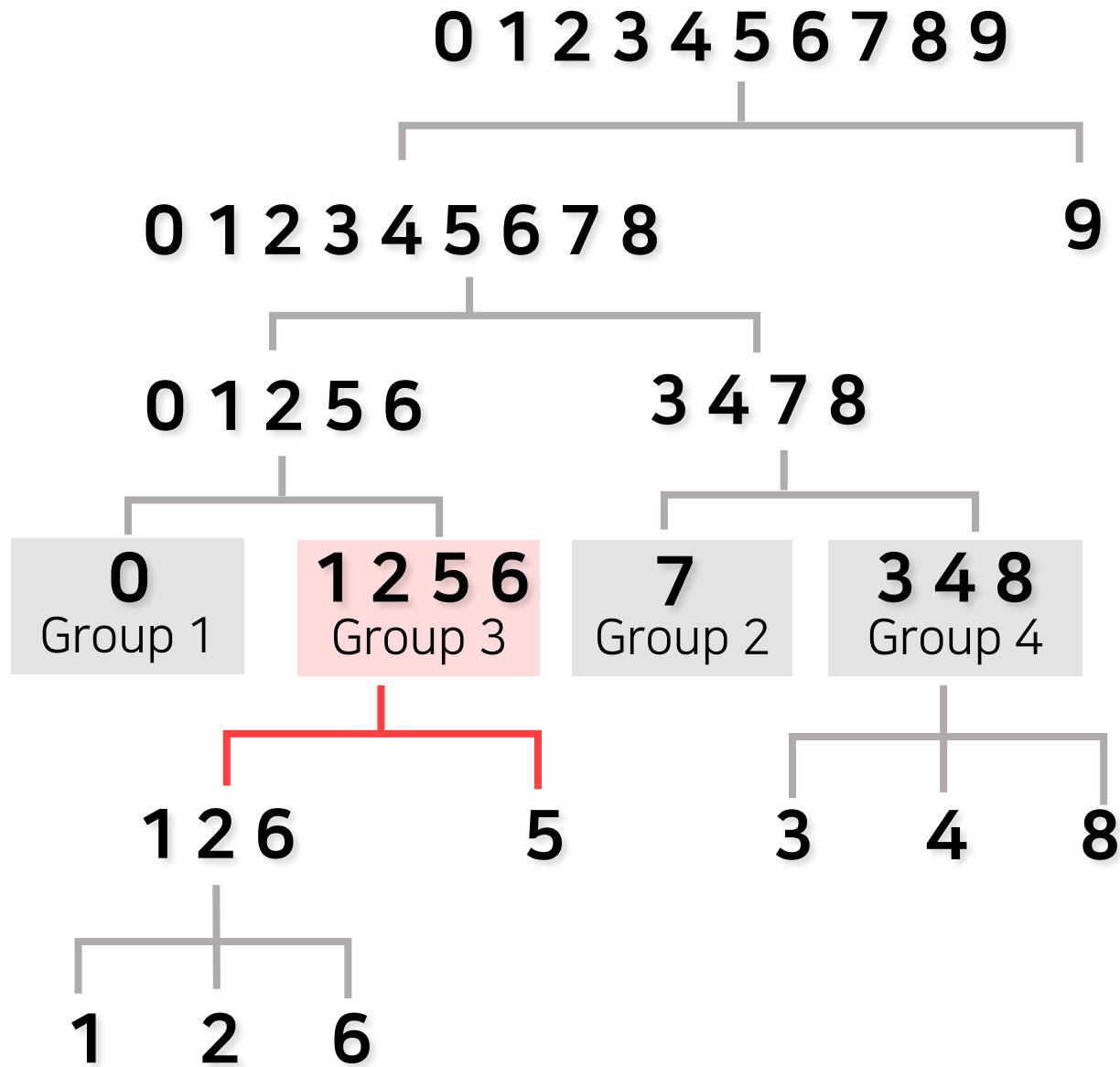
4분위수란 관측값을 작은 순서로 배열했을 때 전체를 사등분하는 값, 극단 값에 영향을 받지 않게 되고, 한쪽으로 치우친 분포에서 극단 값을 제외한 퍼진 정도를 알려고 할 때 사용한다.



제3사분위수, 누적 백분율이 75%에 해당하는 값



단위 : Hz		오류값
Num	Max(F1)	
0	855	676
1	524	
2	284	
3	946	1083
4	969	
5	613	
6	500	885
7	555	
8	1128	
오류값 제거		



○ F2 - F1의 평균 차이  
Group3에서 확연히  
낮은 값을 5로 리턴

○ F2 - F1의 평균 차이를 구할 때 수행착오  
Group3에서 확연히 낮은 값을 5로 리턴

문제

Undefined 값이 발생해 이용하려는 특징인  
F2-F1의 평균값에 영향을 미치는 것을 발견

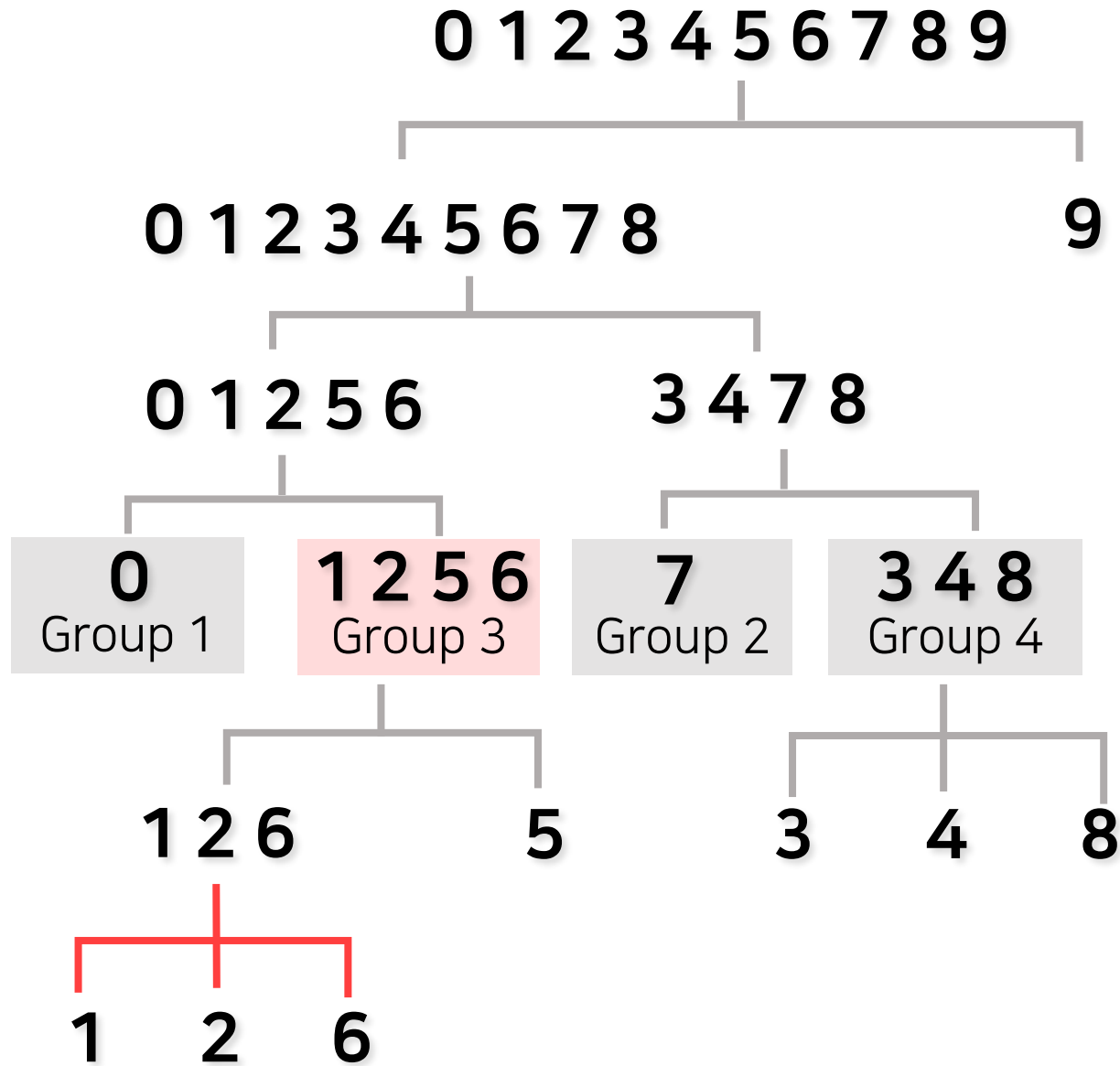
기본 음성 파일에서는 undefined값이 다소 적었으나 음성  
파일의 앞 뒤에 공백이 생길 때는 많은 undefined값 발생

해결책

- 정확도 향상을 위해 피치 구간이 정의된 구간  
에서의 포먼트 값만 사용
- f1, f2, f3, f4 중 하나라도 undefined값이 존  
재하면 사용하지 않음

time	f1	f2	f3	f4
0.334062	556.547033	2010.185040	5448.578357	--undefined--
0.340312	554.228798	2036.185222	5448.576160	--undefined--
0.346562	551.749547	2062.532576	5448.574190	--undefined--
0.352812	549.114188	2088.837466	5448.572667	--undefined--
0.359062	546.423040	2114.011581	5448.572127	--undefined--
0.365312	543.726331	2137.375575	5448.573347	--undefined--
0.371562	541.132198	2158.670776	5448.577845	--undefined--
0.377812	538.655448	2177.704957	5448.588490	--undefined--
0.384062	536.370294	2194.189145	5448.610329	--undefined--
0.390312	534.308488	2208.549340	5448.653061	--undefined--
0.396562	532.553044	2219.724465	5448.733298	--undefined--
0.402812	531.035843	2229.038391	5448.889712	--undefined--
0.409062	529.773447	2236.674235	5449.195292	--undefined--
0.415312	528.756979	2242.939221	5449.804047	--undefined--
0.421562	527.972712	2247.742080	--undefined--	--undefined--
0.427812	527.379556	2251.899675	--undefined--	--undefined--
0.434062	526.955936	2256.360292	--undefined--	--undefined--
0.440312	526.716478	2261.246604	--undefined--	--undefined--
0.446562	526.670713	2267.575789	--undefined--	--undefined--
0.452812	526.832892	2277.028878	--undefined--	--undefined--
0.459062	527.233528	2291.539032	--undefined--	--undefined--
0.465312	527.884665	2313.975762	--undefined--	--undefined--
0.471562	528.706805	2345.838089	--undefined--	--undefined--
0.477812	529.573095	2386.728882	--undefined--	--undefined--

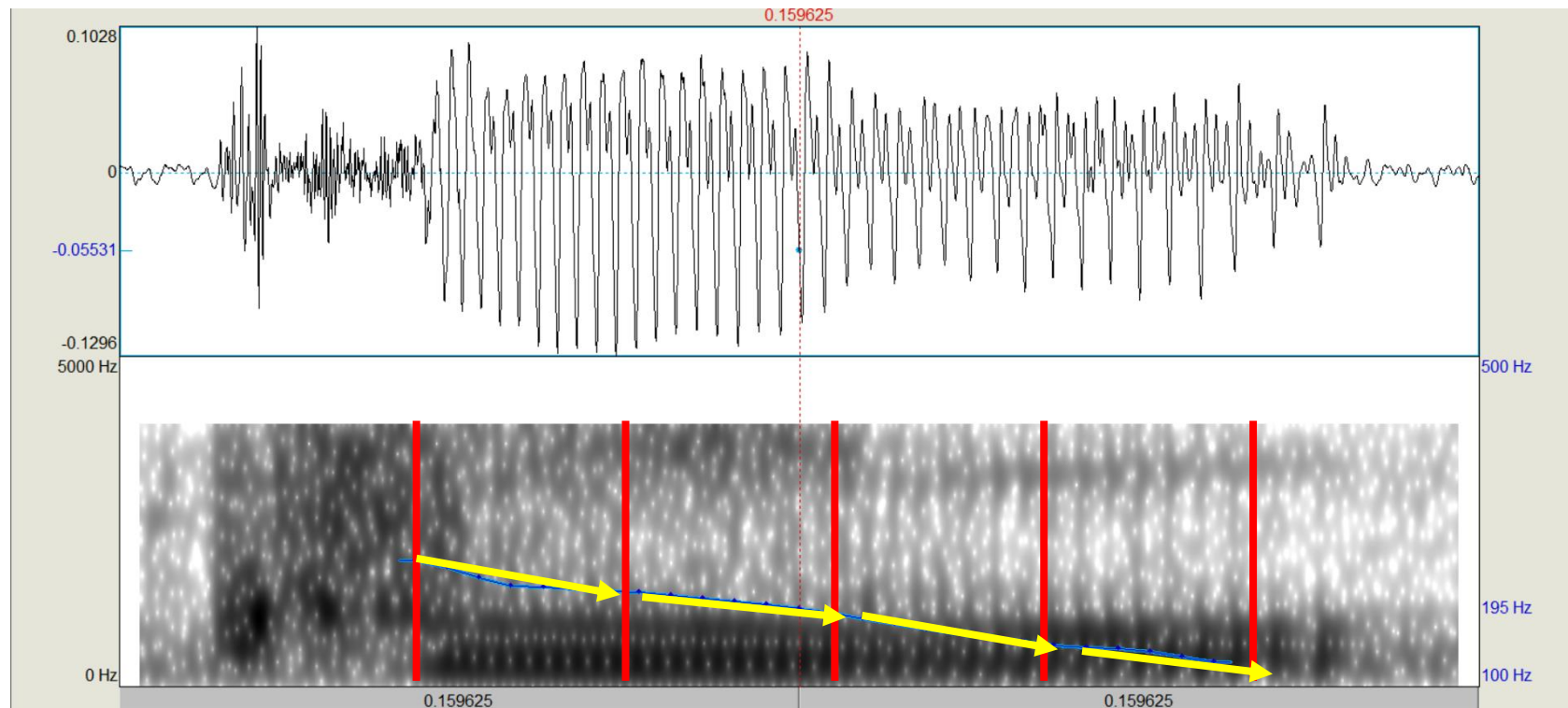
...



○ Picth의 그래프 개형  
변화 관찰

## ○ 5를 제외한 Group4 (1, 2, 6) 구분

Pitch가 생성된 구간을 4등분하여 구간마다  
피치 그래프 개형이 증가/감소 하는지 확인

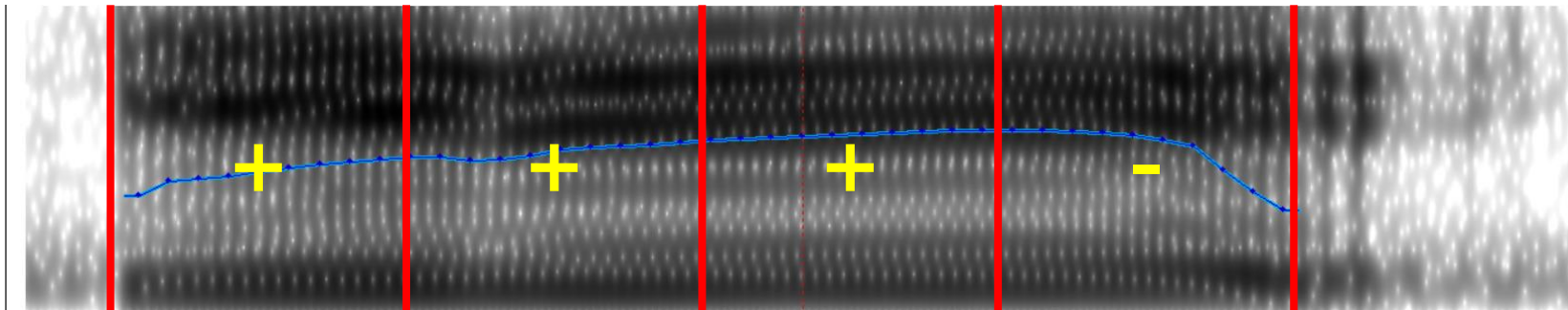




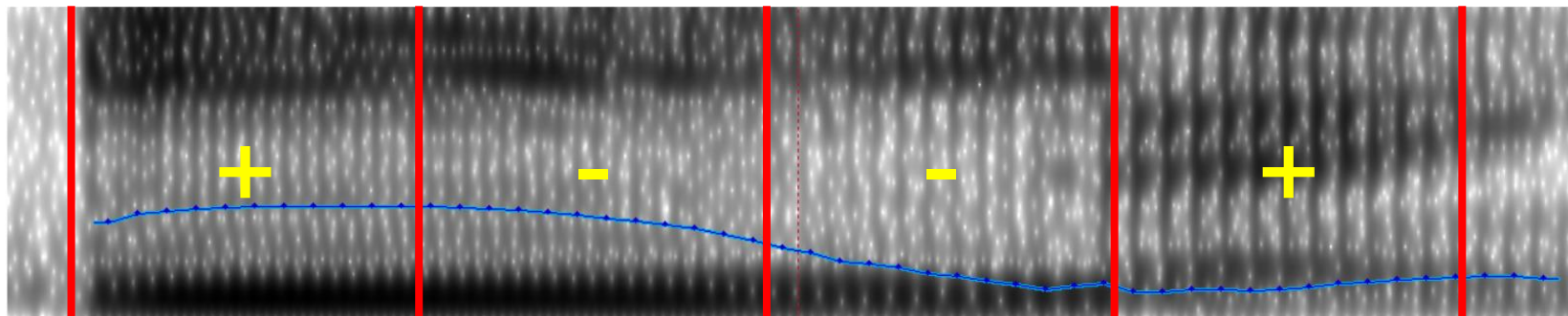
○ 5를 제외한 Group4 (1, 2, 6) 구분

Pitch가 생성된 구간을 4등분하여 구간마다  
피치 그래프 개형이 증가/감소 하는지 확인

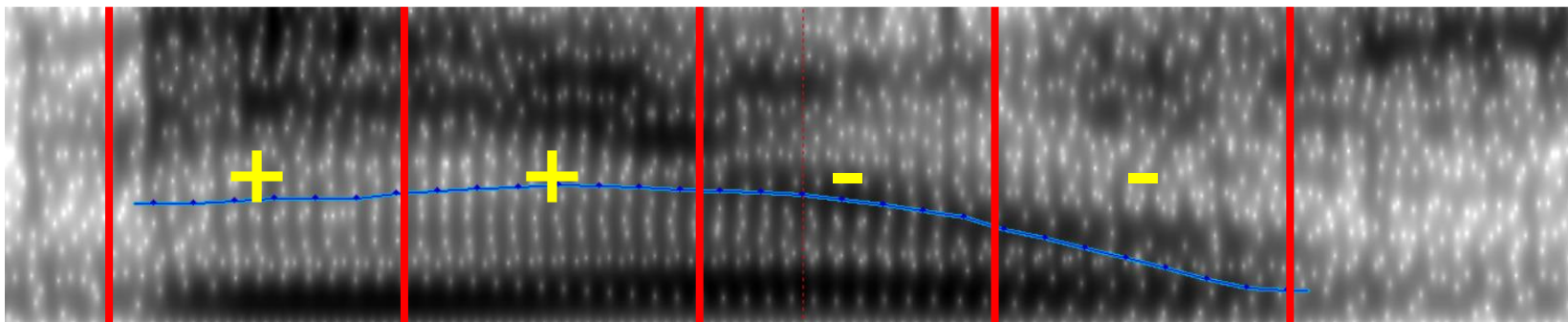
숫자 1



숫자 2



숫자 6

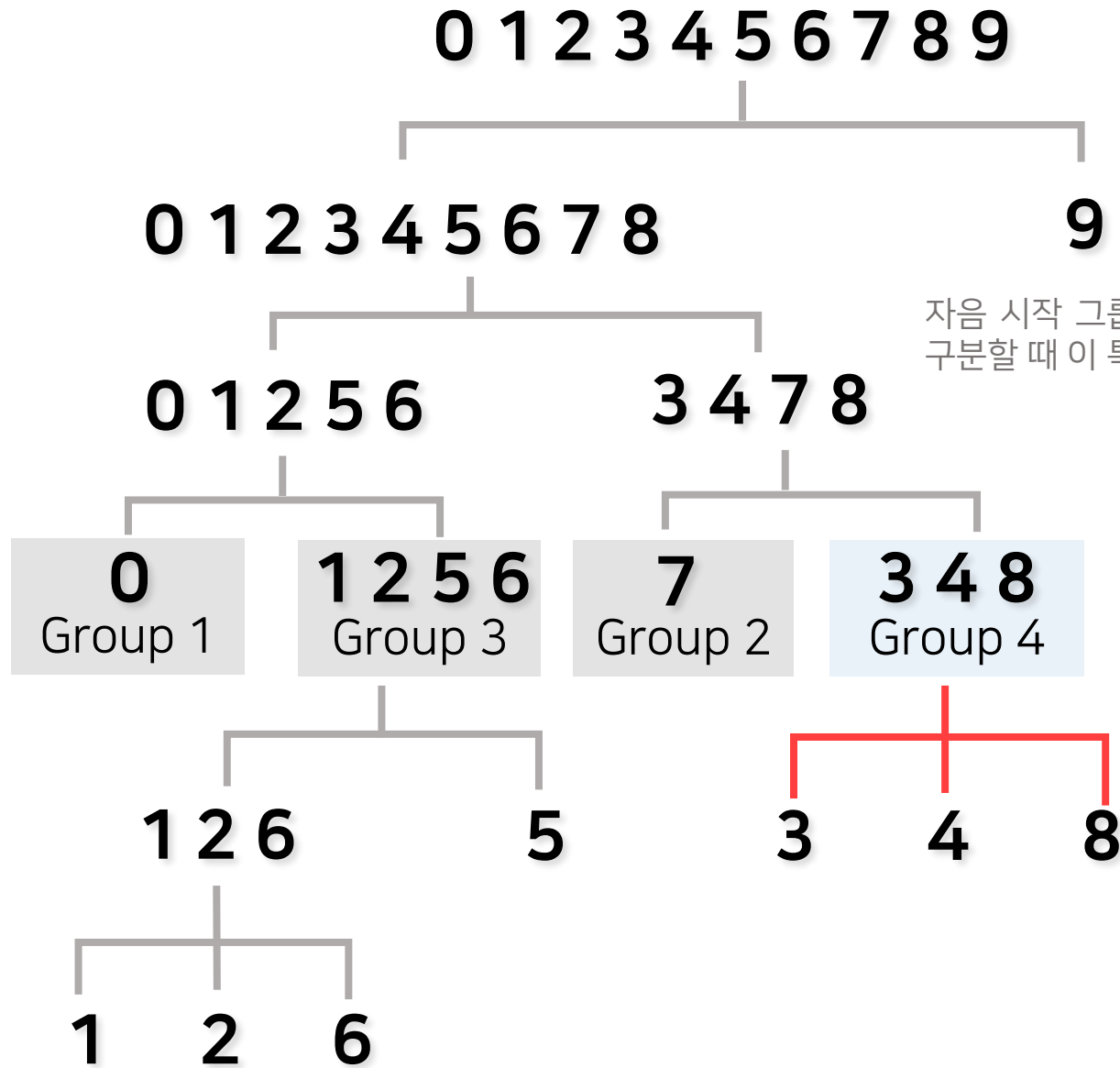




## ○ 5를 제외한 Group4 (1, 2, 6) 구분

Pitch가 생성된 구간을 4등분하여 구간마다  
피치 그래프 개형이 증가/감소 하는지 확인

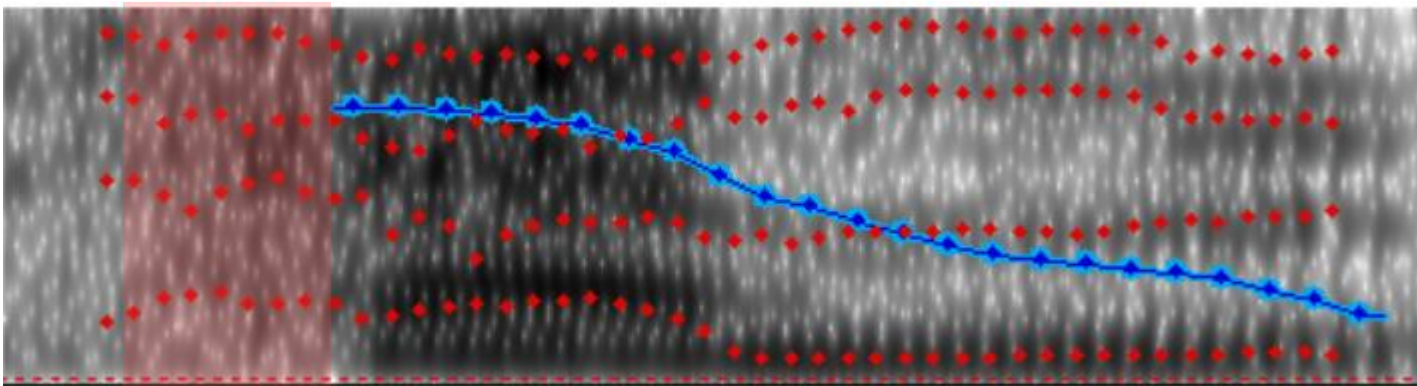
1구간	2구간	3구간	4구간	num
+	+	+	-	1
+	-	-	+	2
+	+	-	-	6



다음 시작 그룹과 모음시작 그룹을  
구분할 때 이 특징을 사용했었음

- Pitch 정의 구간 재사용  
Praat 분석프로그램 상에  
서 Pitch가 정의되는 구간  
과 포먼트가 정의되는 구간  
의 시작점의 차이를 이용함

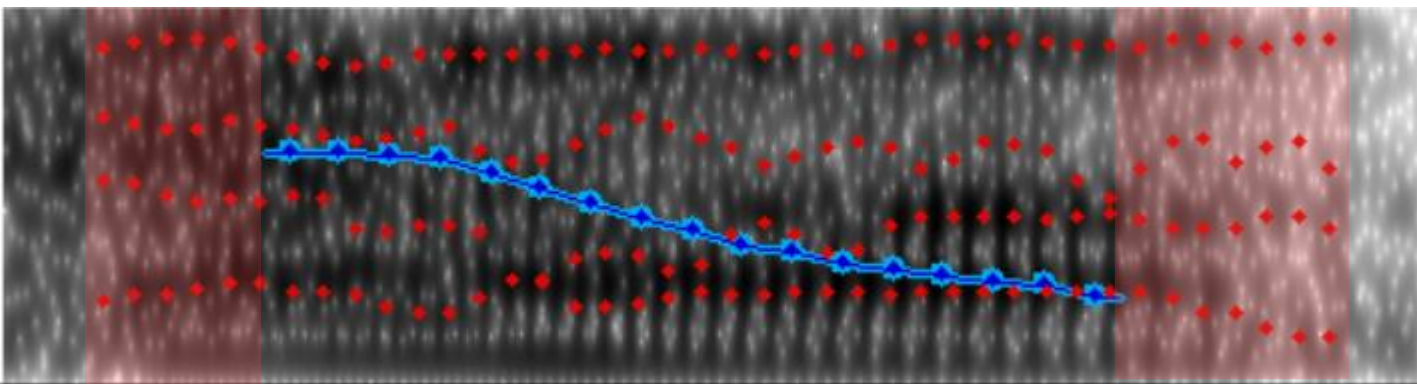
숫자  
3



Pitch 시점 - Formant 시점 절댓값

- 3 0.054 → 5
- 4 0.037 → 3
- 8 0.07 → 7

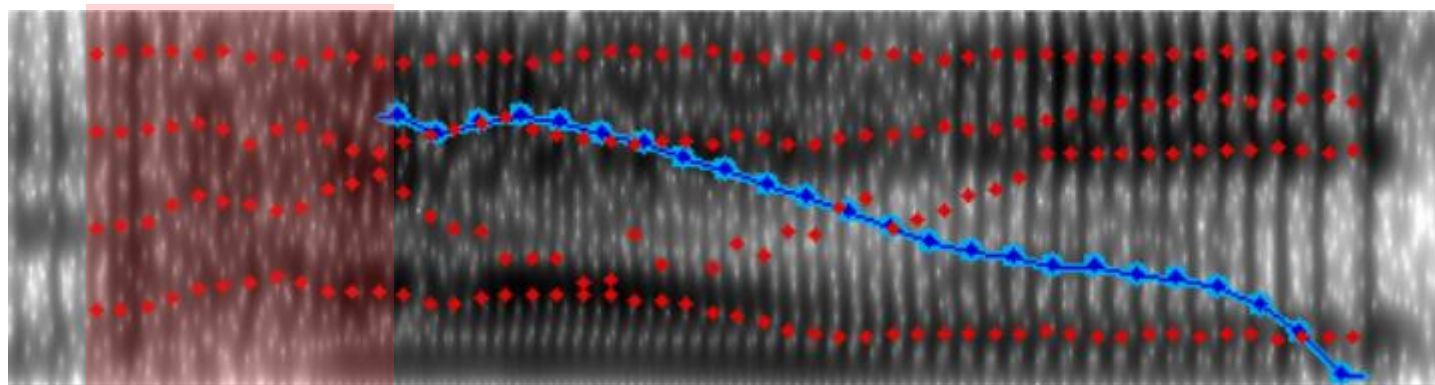
숫자  
4



○ Pitch 정의 구간 재사용  
Praat 분석프로그램 상에서 Pitch가 정의되는 구간과 포먼트가 정의되는 구간의 시작점의 차이를 이용함

3, 4, 8이 피치 구간과 포먼트 구간에서 확실한 차이를 보이기 때문에 구분 가능

숫자  
8



## 테스트 과정

### 테스트 데이터

▶ 볼륨 조절 (50% 200%)

▶ 위치 조절 (앞, 뒤 공백)



0.5



1



2.0

## 테스트 결과 & 문제점 분석

5/10

### 문제점 분석

시작음의 자음 모음 구분 특징

→ 포먼트와 피치 구간의 차이

### 해결방안 및 계획

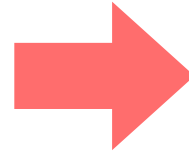
분석 프로그램의 성능으로부터 영향을  
받지않는 특징을 도출

## 계획

분석 프로그램의 성능으로부터 영향을  
받지않는 새로운 특징을 도출

기존의 특징들을 이용한  
각 함수들을 재검토

흐름의 영향을 줄이기 위해  
함수들 개별화 시도



성능 및 인식률  
향상이 기대됨