

개인 프로젝트 자료

목소리로 배란기와 생리예정일을 예측하는 프로젝트

Index

1. 프로젝트 요약

2. 개발 동기

3. 프로젝트 기대효과

4. 핵심 기능

5. 프로젝트 구현

6. 분석

7. 결론

프로젝트 요약

구분		2018												2019		
		1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
데이터 수집환경 조성	목소리를 분석해 수집할 데이터 셋 도출															
	음성학 및 음성분석 툴 ‘Praat’ 학습															
	목소리 녹음 앱 개발															
목소리 녹음	1차 녹음 (5명)															
	녹음 환경 수정 및 음성 데이터 정규화															
	2차 녹음 (14명)															
	3차 녹음 (1명)															
생리예정일 예측	지도 학습(머신러닝) 학습															
	학습 데이터 및 테스트 데이터 수집															
	데이터 전처리															
	다중 선형회귀모델 생성 및 적용															
	프로토타입 자동화															

프로젝트 내용	- 목소리를 녹음하면 생리예정일이 언제인지 예측해주는 프로젝트
프로젝트 결과	<div> <div>- 프로젝트를 통해 7일의 오차가 있던 3차 녹음 여성의 생리예정일을 2일 이내의 오차로 예측</div> <div>- 2,3차 녹음 여성 모두 생리주기동안 유사한 특징을 보임</div> <div>(배란기 때는 목소리가 맑아지고, 생리 때는 목소리가 거칠어진다. 그리고 그 차이는 유의미)</div> </div>

개발 동기

목소리와 배란기의 상관관계를 실험한 논문을 보고 영감

논문 제목: 여성의 목소리가 배란기를 예측하는 데에 단서를 제공하는가?

논문 저자: Cognitive Ethology Laboratory, German Primate Center, Germany

Do Women's Voices Provide Cues of the Likelihood of Ovulation? The Importance of Sampling Regime

Julia Fischer^{1*}, Stuart Semple², Gisela Fickenscher¹, Rebecca Jürgens¹, Eberhard Kruse³, Michael Heistermann⁴, Ofer Amir⁵

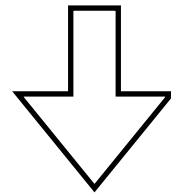
¹ Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany, ² Centre for Research in Evolutionary Anthropology, Roehampton University, London, United Kingdom, ³ Department of Phoniatrics and Pedaudiology, Georg-August Universität Göttingen, Göttingen, Germany, ⁴ Reproductive Biology Unit, German Primate Center, Göttingen, Germany, ⁵ Department of Communication Disorders, Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

Abstract

The human voice provides a rich source of information about individual attributes such as body size, developmental stability and emotional state. Moreover, there is evidence that female voice characteristics change across the menstrual cycle. A previous study reported that women speak with higher fundamental frequency (F0) in the high-fertility compared to the low-fertility phase. To gain further insights into the mechanisms underlying this variation in perceived attractiveness and the relationship between vocal quality and the timing of ovulation, we combined hormone measurements and acoustic analyses, to characterize voice changes on a day-to-day basis throughout the menstrual cycle. Voice characteristics were measured from free speech as well as sustained vowels. In addition, we asked men to rate vocal attractiveness from selected samples. The free speech samples revealed marginally significant variation in F0 with an increase prior to and a distinct drop during ovulation. Overall variation throughout the cycle, however, precluded unequivocal identification of the period with the highest conception risk. The analysis of vowel samples revealed a significant increase in degree of unvoiceness and noise-to-harmonic ratio during menstruation, possibly related to an increase in tissue water content. Neither estrogen nor progesterone levels predicted the observed changes in acoustic characteristics. The perceptual experiments revealed a preference by males for voice samples recorded during the pre-ovulatory period compared to other periods in the cycle. While overall we confirm earlier findings in that women speak with a higher and more variable fundamental frequency just prior to ovulation, the present study highlights the importance of taking the full range of variation into account before drawing conclusions about the value of these cues for the detection of ovulation.

개발 동기

증가하는 불임 부부
비싼 산부인과 진료비
생리불순에 대한 관리 부실



배란기를 간편하면서도 정확하게 파악한다면 문제를 해결할 수 있다고 생각
핸드폰으로 녹음만 한다면 배란기와 생리예정일을 알 수 있는 앱을 만들고자 함

프로젝트 기대 효과



배란 테스트기 및 임신 테스트기 대체



경구 피임약 효과 알리미

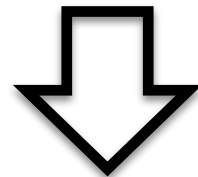
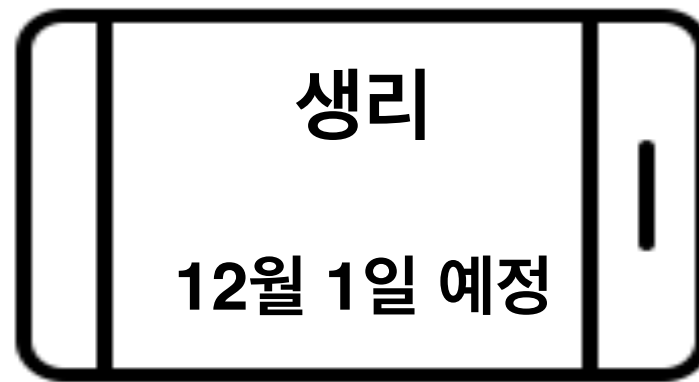
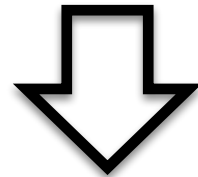


여성 건강 알리미

배란 테스트기 및 임신 테스트기 대체	<ul style="list-style-type: none"> -정자는 최대 5일, 난자는 최대 6시간 생존 -난자가 나오는 시기(배란기)만 알 수 있다면 임신, 피임을 하는데에 큰 도움이 됨
경구 피임약 효과 알리미	<ul style="list-style-type: none"> -현재 경구 피임약을 제대로 먹고 있는지 알 수 있는 방법 부재 -본 프로젝트를 통해 경구 피임약을 제대로 먹고 있는지 확인할 수 있음
여성 건강 알리미	<ul style="list-style-type: none"> -정상 생리주기를 가진 여성의 음성 데이터를 학습 -사용자의 생리주기에 이상 징후가 발견되면 이에 대한 정보를 알림

핵심 기능

목소리 녹음



배란 예정일 정보 제공

*여성 호르몬의 변화에 따라 목소리의 거침정도가 달라지는 원리를 이용

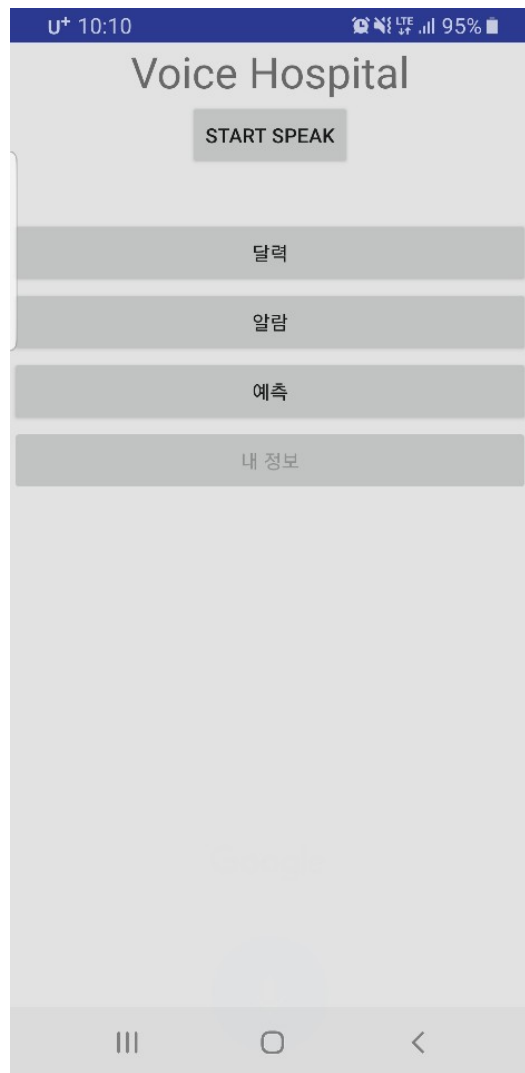
*목소리 녹음을 많이할수록 정확도가 높아짐

핵심 기능

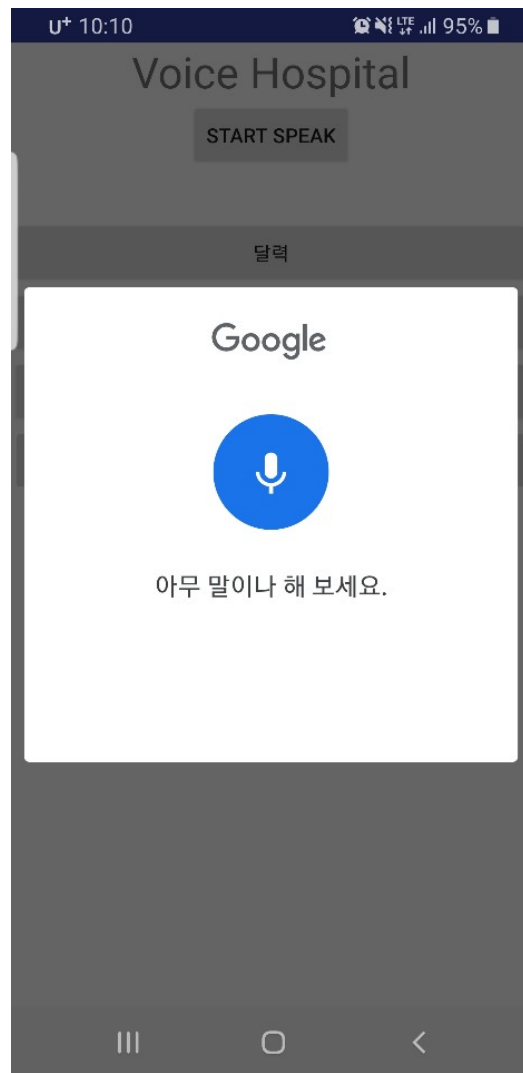
안드로이드 앱	음성 분석 툴 ‘Praat’	Tensorflow
<div><div>주 기능</div><div>‘파크 파크 파크 파크 파크’ 목소리 녹음</div><div>↓</div><div>단어가 틀리다면, 오류 메시지 전송 단어가 맞으면 Filezilla에 음성파일 업로드</div></div>	<div>음성 파일로부터 4가지 변수 추출 (DVB, DUV, NHR, MEAN PITCH) ↓ csv 파일로 정리</div>	<div>학습시킨 회귀모델을 통해 녹음 파일 생리예정일 예측</div>
<div><div>부 기능</div><div>생리달력으로 생리 데이터 입력 알람기능으로 사용자에게 알람 기능 제공</div></div>		

DVB: 목소리의 변성 정도
DUV: 무성음의 정도
NHR(음향 변수): 음파가 전파하면서 양적으로 변화를 일으키는 것
MEAN PITCH: 음높이 평균

프로젝트 구현



메인 화면



목소리 녹음
(음성 인식 기능 탑재)

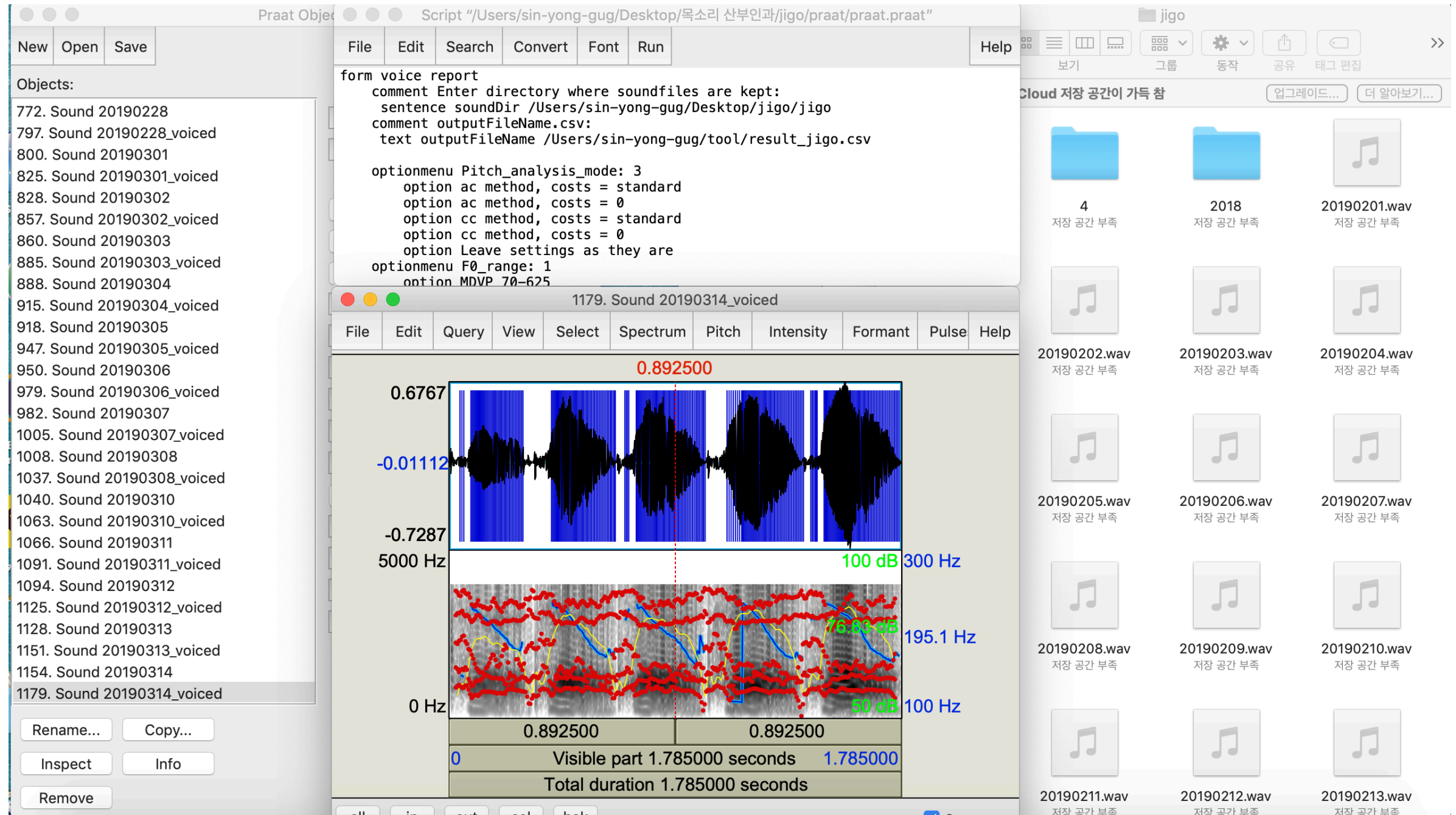


목소리 녹음 시간 알람



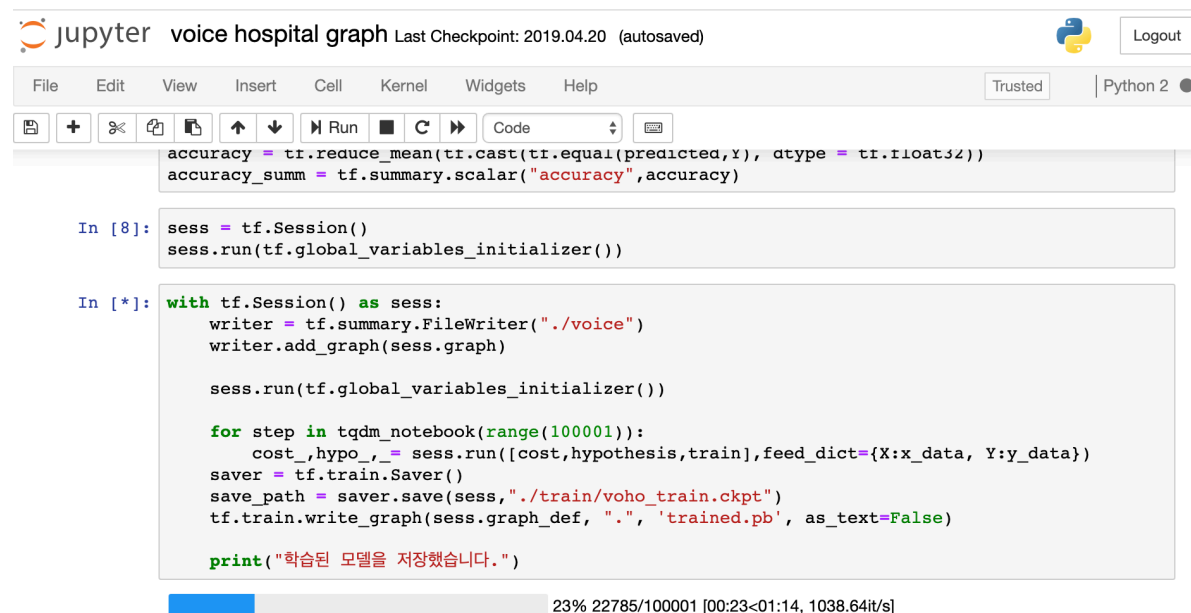
생리 달력

프로젝트 구현



음성분석 툴 'Praat'를 통한 음성분석

프로젝트 구현



The image shows a Jupyter Notebook interface with the title "voice hospital graph". The top bar indicates the last checkpoint was saved on 2019.04.20. The notebook contains three code cells. The first cell defines accuracy calculation. The second cell initializes a TensorFlow session. The third cell is a training loop that runs for 100,001 steps, saving the model as "trained.pb".

```
accuracy = tf.reduce_mean(tf.cast(tf.equal(predicted, y), dtype=tf.float32))
accuracy_summ = tf.summary.scalar("accuracy", accuracy)

In [8]: sess = tf.Session()
sess.run(tf.global_variables_initializer())

In [*]: with tf.Session() as sess:
    writer = tf.summary.FileWriter("./voice")
    writer.add_graph(sess.graph)

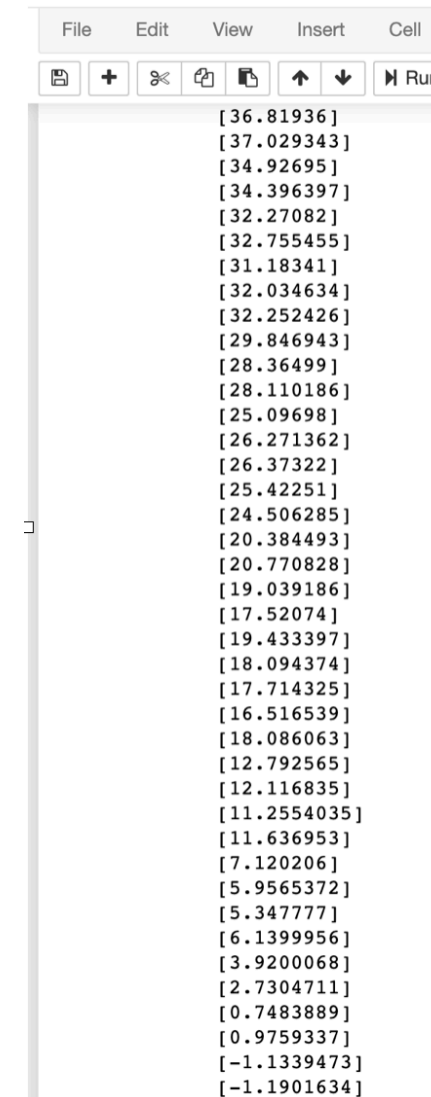
    sess.run(tf.global_variables_initializer())

    for step in tqdm_notebook(range(100001)):
        cost, hypo_, _ = sess.run([cost, hypothesis, train], feed_dict={X:x_data, Y:y_data})
    saver = tf.train.Saver()
    save_path = saver.save(sess, "./train/voho_train.ckpt")
    tf.train.write_graph(sess.graph_def, ".", 'trained.pb', as_text=False)

    print("학습된 모델을 저장했습니다.")
```

23% 22785/100001 [00:23<01:14, 1038.64it/s]

음성 분석 데이터를 통한
다중 선형회귀 모델 학습



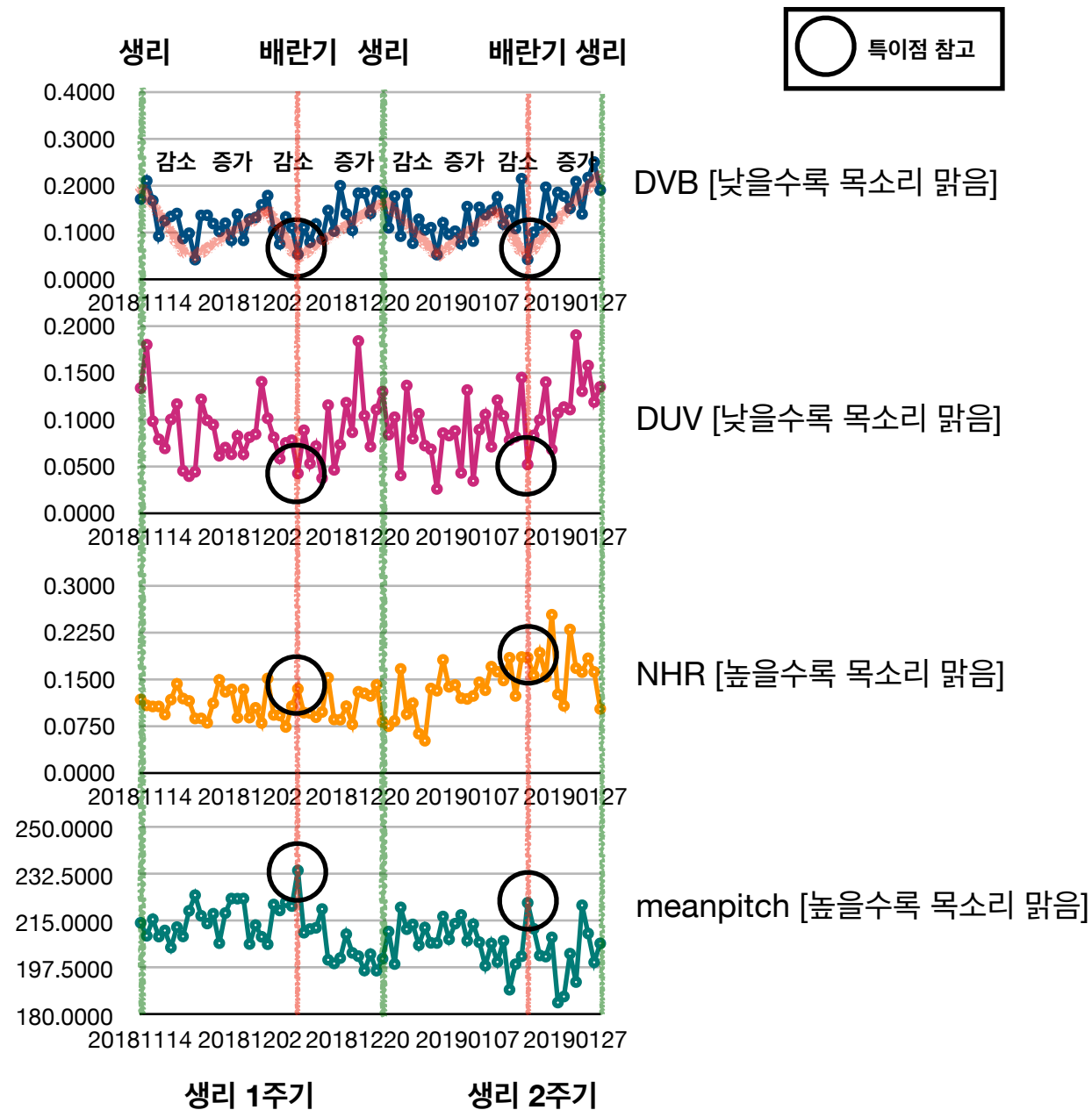
The image shows a Jupyter Notebook interface with a single code cell containing a list of 30 numerical values, likely representing model predictions or costs.

```
[36.81936]
[37.029343]
[34.92695]
[34.396397]
[32.27082]
[32.755455]
[31.18341]
[32.034634]
[32.252426]
[29.846943]
[28.36499]
[28.110186]
[25.09698]
[26.271362]
[26.37322]
[25.42251]
[24.506285]
[20.384493]
[20.770828]
[19.039186]
[17.52074]
[19.433397]
[18.094374]
[17.714325]
[16.516539]
[18.086063]
[12.792565]
[12.116835]
[11.2554035]
[11.636953]
[7.120206]
[5.9565372]
[5.347777]
[6.1399956]
[3.9200068]
[2.7304711]
[0.7483889]
[0.9759337]
[-1.1339473]
[-1.1901634]
```

학습 모델을 활용한 생리 예정일 예측

분석

가설 : 배란기 때 목소리가 맑아지고, 생리 때 목소리가 거칠어진다.



확인사항	결과
가설과의 부합 여부	테스팅 한 여성 모두 배란기 때 목소리가 맑아지고 생리 때 목소리가 거칠어졌음
특징	<ul style="list-style-type: none"> - DVB에서 W 모양을 그림 - meanpitch에서 가장 높은 값이 배란기 - DUV에서 가장 낮은 값이 배란기 - 생리기간 ~ 배란기까지 NHR값 증가 유무
반복 유무	4가지 변수가 보인 특징이 반복해서 나타남

* 배란기: 난자가 배출하는 여성의 월경 주기의 한 과정

분석

Coefficients :

	Estimate	Std. Error	t	value	Pr(> t)	
(Intercept)	4.012E+01	2.755E-02	145	5.906	< 2e-16	***
day	-5.013E-02	5.087E-05	-98	5.497	< 2e-16	***
diff	-9.501E-01	8.396E-05	-1131	5.381	< 2e-16	***
dvb	-1.202E-01	3.581E-02	-	3.356	0.00128	**
duv	7.433E-02	3.994E-02		1.861	0.06692	.
nhr	4.018E-02	2.573E-02		1.562	0.12289	
mean	-2.733E-04	1.207E-04	-	2.263	0.02674	*

Signif. code s: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error:	0.007203	on 70 degrees of freedom
Multiple R-squared:	1,	Adjusted R-squared: 1
F-statistic:	3.17e+07 on 6 and 70 DF,	p-value: < 2.2E-16

학습데이터 신뢰도

본인의 데이터로 학습시킨 모델을 통해
다음 생리주기를 예측한 결과

P-value가 2.2e-16 이하로
학습데이터가 생리 예정일과 매우 유의미한 상관관계가 있음을 나타낸다.

실제 생리 예정일 남은 일 수			예측한 생리 예정일 남은 일 수		
.....			실제	예측	차이
37	37	0	19	18	1
36	34	2	18	18	0
35	36	-1	17	14	3
34	34	0	16	15	1
33	33	0	15	14	1
32	31	1	14	14	0
31	31	0	13	14	-1
30	29	1	12	12	0
29	29	0	11	11	0
28	29	-1	10	11	-1
27	27	0	9	6	3
26	27	-1	8	5	3
25	27	-2	7	7	0
24	24	0	5	3	2
23	24	-1	4	7	-3
22	22	0	3	5	-2
21	19	2	1	2	-1
20	20	0	0	0	0

사용하고 있던 생리 달력 앱에서는 13일의 오차로 생리예정일을 예측했지만,
프로젝트를 통해 2일 이내의 오차로 생리예정일 예측