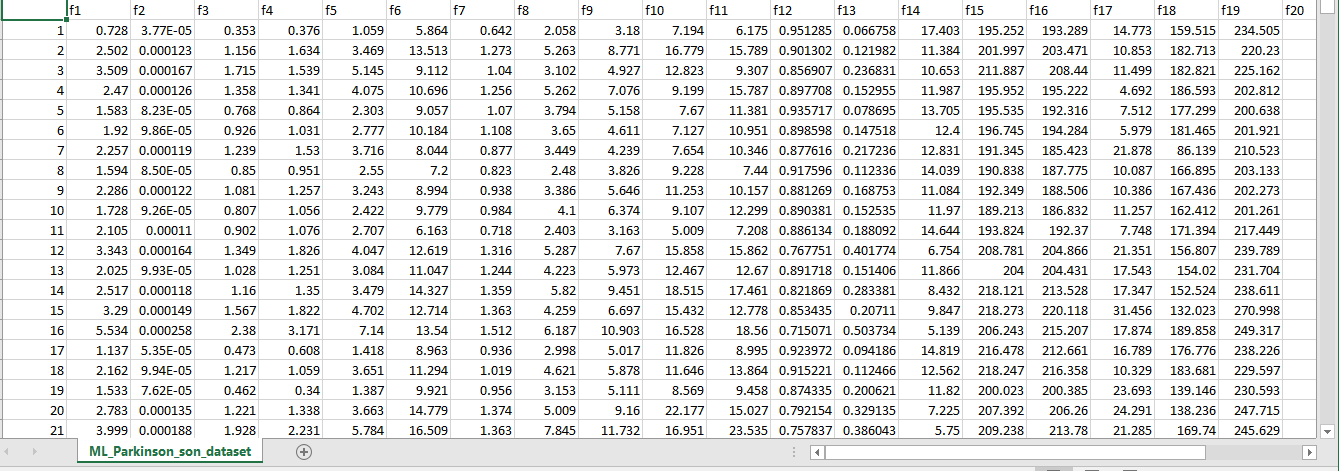
**UCI Parkinson’s speech data**

The dataset was developed by the Dept. of Neurology in the cerrahpasa Faculty of Medicine, Istanbul University [1]. This dataset consists of 23 voice sampled sound recordings captures the sustainable sounds of numbers, words, Vowels and short sentences collected from the 20 healthy patients, 20 from PD Patients. The linear and time frequency-based features are extracted from the sound samples. The datasets have 1040 instances. The ‘Class’ attribute in the dataset differentiate the healthy and PD Patients; 0 states that the person is healthy, while 1 state that the person has Parkinson’s disease. Fig 1 illustrates the sample of the data set used. In the research Methodological design, the 19 voice features are selected and used with the 1040 instances.



### Figure 1: Dataset of biomedical Voice measurements of 31 people

### Table 1: Extracted Features from Speech Recordings

|  |  |
| --- | --- |
| Feature on Columns | Features |
| F1  F2  F3  F4  F5 | Jitter (local)  Jitter (local, absolute)  Jitter (rap)  Jitter (ppq5)  Jitter (ddp) |
| F6  F7  F8  F9  F10  F11 | Shimmer(local)  Shimmer(local,dB)  Shimmer(apq3)  Shimmer(apq5)  Shimmer(apq11)  Shimmer(dda) |
| F12  F13  F14 | AC  NTH  HTN |
| F15  F16  F17  F18  F19 | Median Pitch  Mean Pitch  Standard Deviation  Maximum Pitch  Minimum Pitch |
| F20  F21  F22  F23 | Number of cycles  Phase Value  Frequency Value  Standard Deviation Period |

1. Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O., 'Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings', IEEE Journal of Biomedical and Health Informatics, vol. 17(4), pp. 828-834, 2013.