**Why Voice Recordings and Parkinson's Disease?**

Parkinson's disease affects the muscles and nerves that control speech. People with Parkinson's often experience changes in their voice, such as:

* **Softer speech** (reduced loudness)
* **Monotone voice** (less variation in pitch)
* **Tremors or shakiness** in the voice
* **Hoarseness** or breathiness

These changes happen because Parkinson's affects the brain's ability to control the muscles used for speaking. By analyzing voice recordings, we can detect these subtle changes and use them to identify whether someone might have Parkinson's.

**Example: How It Works**

Imagine two voice recordings:

1. **Healthy Person**:
   * Smooth pitch (low jitter).
   * Stable loudness (low shimmer).
   * Clear voice (high HNR, low NHR).
2. **Person with Parkinson's**:
   * Shaky pitch (high jitter).
   * Unstable loudness (high shimmer).
   * Breathy or hoarse voice (low HNR, high NHR).

The dataset uses these differences to help identify Parkinson's disease.

**What Does "MDVP" Mean?**

MDVP stands for **Multi-Dimensional Voice Program**. It's a software tool used to analyze voice recordings and extract features like pitch, loudness, and noise. The "Hz" (Hertz) is the unit for frequency, which is used to measure pitch.

**Summary**

* The dataset contains **voice features** like pitch, loudness, and noise.
* These features help detect **changes in the voice** caused by Parkinson's disease.
* The goal is to use these features to **predict whether someone has Parkinson's**.

**1. name**

* This is just the **ID or name** of the voice recording. It helps us keep track of which recording we're looking at.

**2. MDVP:Fo (Hz)**

* This is the **average pitch** of the voice. Think of it as how high or low someone's voice sounds on average.

**3. MDVP:Fhi (Hz)**

* This is the **highest pitch** in the voice recording. It tells us the top note the person reached while speaking.

**4. MDVP:Flo (Hz)**

* This is the **lowest pitch** in the voice recording. It tells us the lowest note the person reached while speaking.

**5. MDVP:Jitter (%)**

* This measures how **shaky** or unstable the voice is. If someone's voice trembles a lot, this value will be higher.

**6. MDVP:Jitter (Abs)**

* This is another way to measure how shaky the voice is, but in **absolute numbers** instead of percentages.

**7. MDVP:RAP**

* This measures how much the **pitch changes** over short periods. A higher value means the pitch is less stable.

**8. MDVP:PPQ**

* This is similar to RAP but measures pitch changes in a slightly different way.

**9. Jitter:DDP**

* This is another measure of how much the **pitch varies**. It’s calculated by looking at the differences between consecutive pitch changes.

**10. MDVP:Shimmer**

* This measures how much the **loudness** of the voice varies. If someone's voice gets louder and softer unexpectedly, this value will be higher.

**11. MDVP:Shimmer (dB)**

* This is the same as Shimmer, but measured in **decibels** (a unit for loudness).

**12. Shimmer:APQ3 , APQ5, APQ**

* These are more detailed ways to measure how much the **loudness varies** over short periods (3, 5, or all periods).

**13. Shimmer:DDA**

* This is another way to measure how much the **loudness varies**, calculated by looking at the differences between consecutive loudness changes.

**14. NHR (Noise-to-Harmonics Ratio)**

* This measures how much **noise** (like breathiness or hoarseness) is in the voice compared to the clear, harmonic parts. A higher value means the voice is noisier.

**15. HNR (Harmonics-to-Noise Ratio)**

* This is the opposite of NHR. It measures how **clear** the voice is. A higher value means the voice is clearer.

**16. status**

* This is the **target column**. It tells us whether the person has Parkinson's disease:
  + **1** = Has Parkinson's.
  + **0** = Does not have Parkinson's.

**17. RPDE**

* This measures the **complexity** of the voice. A higher value means the voice has more complex patterns.

**18. DFA**

* This measures how **smooth** or **rough** the voice is. A higher value means the voice is smoother.

**19. spread1, spread2**

* These measure how much the **pitch varies** in a more advanced way. They help detect subtle changes in the voice.

**20. D2**

* This measures the **complexity** of the voice in another way. A higher value means the voice is more complex.

**21. PPE**

* This measures how **unpredictable** the pitch of the voice is. A higher value means the pitch is less predictable.

**Summary of What Each Column Represents**

* **Pitch-related columns**: Tell us how high, low, or shaky the voice is.
* **Loudness-related columns**: Tell us how stable or unstable the loudness of the voice is.
* **Noise-related columns**: Tell us how clear or noisy the voice is.
* **Complexity-related columns**: Tell us how complex or smooth the voice patterns are.
* **status**: Tells us whether the person has Parkinson's or not.

**Why Are These Features Important?**

People with Parkinson's disease often have:

* A **shaky voice** (high jitter).
* **Unstable loudness** (high shimmer).
* A **breathy or hoarse** voice (high NHR, low HNR).
* Less variation in pitch and loudness (measured by RPDE, DFA, etc.).

By analyzing these features, we can detect Parkinson's disease early using just voice recordings!