Formant Distribution

Distributions of steady state Formants

The main purpose of this demo notebook is to explore formant statistics based on the Hillenbrand database (1995). Hillenbrand wanted to redo the seminal Peterson and Barney experiments from the early 1950's, of which the original speech data was not preserved. Hillenbrand found significant differences from the P-B data, that might be due to a number of factors:
- formant measurements were now based on LPC analysis and not purely human reading from wideband spectrograms
- population choice
- dialectic evolution
- ...

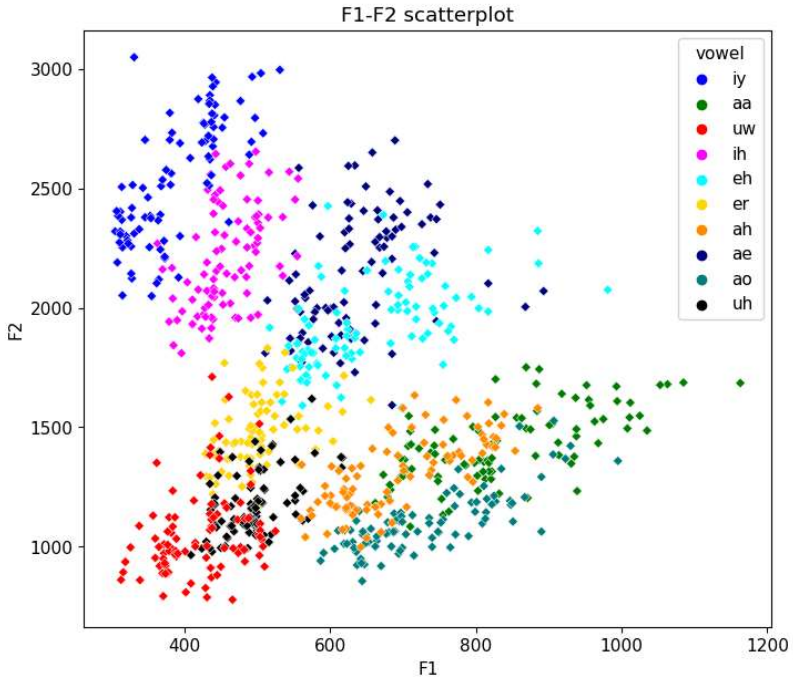
Since these P-B experiments we know that formants (especially F1 and F2) carry great discriminative power when it comes to nizing vowels. In the very early days of automatic speech recognition formant extraction followed by formant based recog ecognition was considered one of the ways to go. This is long obsolete by now. Formants are too ill-defined and as a feature set it is too minimal.
At the same time formants are robust against all kinds of signal manipulations and are illustrative for the tremendous aundancy present in speech (from a recognition point of view).

In this notebook we focus on the 'explorative phase' in which we explore the potential of formants for speech recognition.

The goal is to observe both WITHIN and BETWEEN class differences. Also observe that there are significant side factors entering into this recognition game: gender, age, ..

2. F1-F2 Scatter Plots

Scatter plots are give you an intuitive feeling of how classes differ from one another in "feature space", i.e. for the features that you have selected. We can only visualize this well for 2 dimensions and to some extent for 3D as well.
The scatter plots below are shown for 3, 6, respectively 10 classes (the vowel classes used by Peterson & Barney, 1952). While the 3 classes are perfectly separable in the F1-F2 space, this is only partially true for the 6 classes and not at all anymore for the 10 class data.

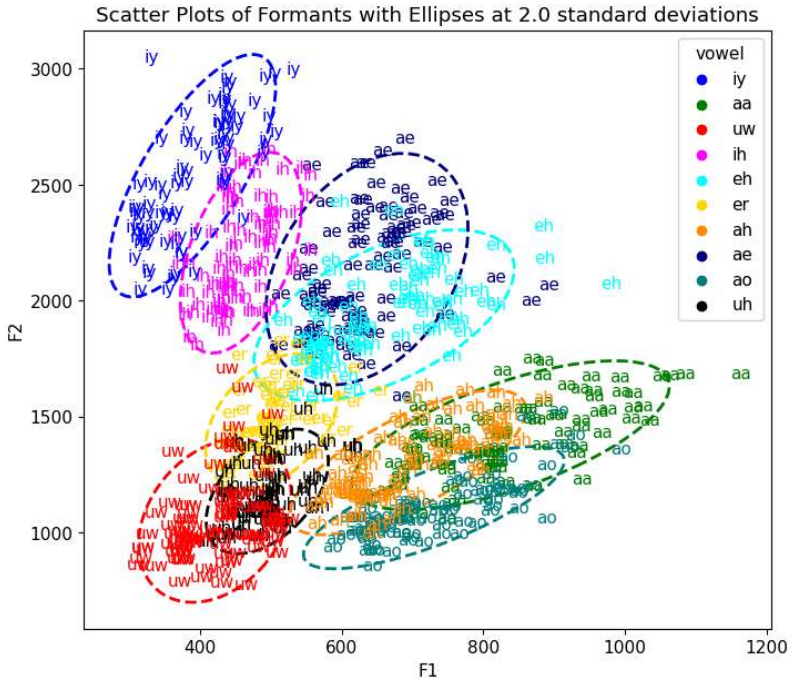


3. Overlaying Scatter plots with Confidence Ellipses

Scatter plots are a non parametric data model.
For pattern recognition purposes we tend to build parametric models that can generalize from the data. Such models may answer better how intrinsically separable the classes are. Here we are using simple gaussian fits and in the plots below we draw confidence ellipses .

Confidence Ellipses using Full Covariance Matrix

In order to plot these confidence ellipses we both measure standard deviation and correlation between the features



4. Formant Tables

First we compute the mean values for the different formants.
We do this for both global means and means per gender.
The table below shows formant values per gender and also gender independent.

Formant Table (gender dependent)

	vowel	iy	ih	eh	ae	aa	ao	uh	uw	ah	er
	gender										
F1	**m**	340	429	588	591	756	656	469	380	621	476
	w	435	484	727	678	916	801	519	460	760	527
F2	**m**	2312	2034	1803	1930	1309	1023	1123	992	1181	1370
	w	2756	2369	2063	2332	1526	1188	1229	1106	1416	1589
F3	**m**	3001	2687	2604	2595	2535	2521	2435	2355	2548	1711
	w	3373	3057	2953	2973	2823	2819	2829	2735	2901	1930

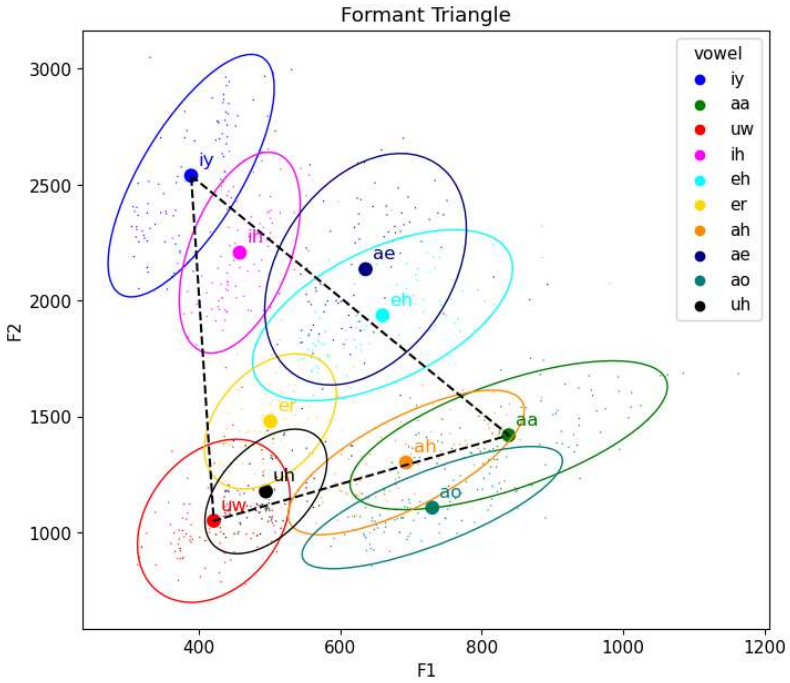
Formant Table (gender independent)

vowel	iy	ih	eh	ae	aa	ao	uh	uw	ah	er
F1	389	458	660	636	838	730	495	421	693	501
F2	2539	2207	1937	2136	1420	1108	1177	1051	1302	1480
F3	3191	2878	2784	2788	2682	2673	2638	2551	2730	1820

5. Formant Triangle

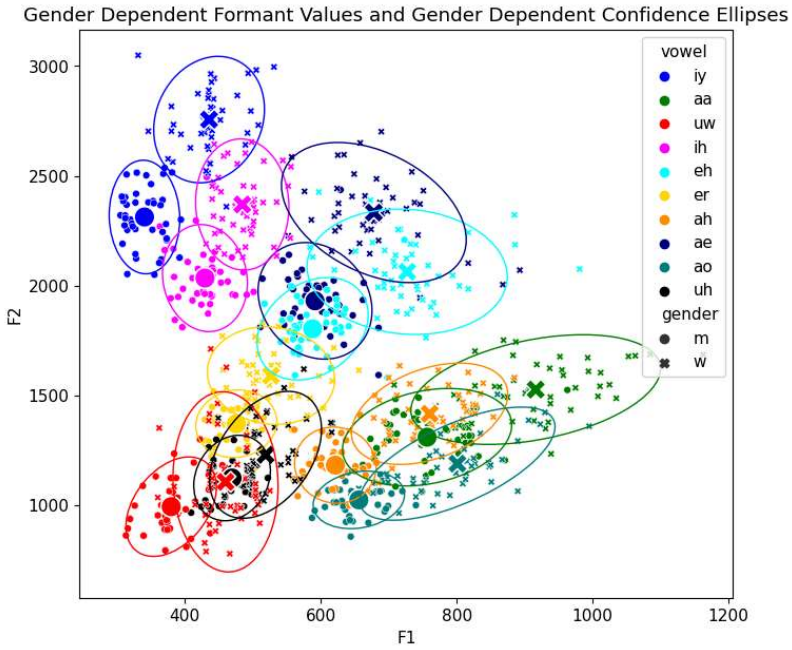
Thte "Formant Triangle" ref[illegible] triangle with as edge points iy, aa and uw.
Remark also that there is one vowel ('er') that doesn't fit the above at all as it is situated in the center of the triangle.

Out[28]:



Gender Dependency of formants

The formant confidence ellipses look quite different when we separate the data per gender. Confusion is obviously much less if the gender would have been known a priori or if it can be inferred from other speech characteristics.



6. Grid plots for higher dimensional data

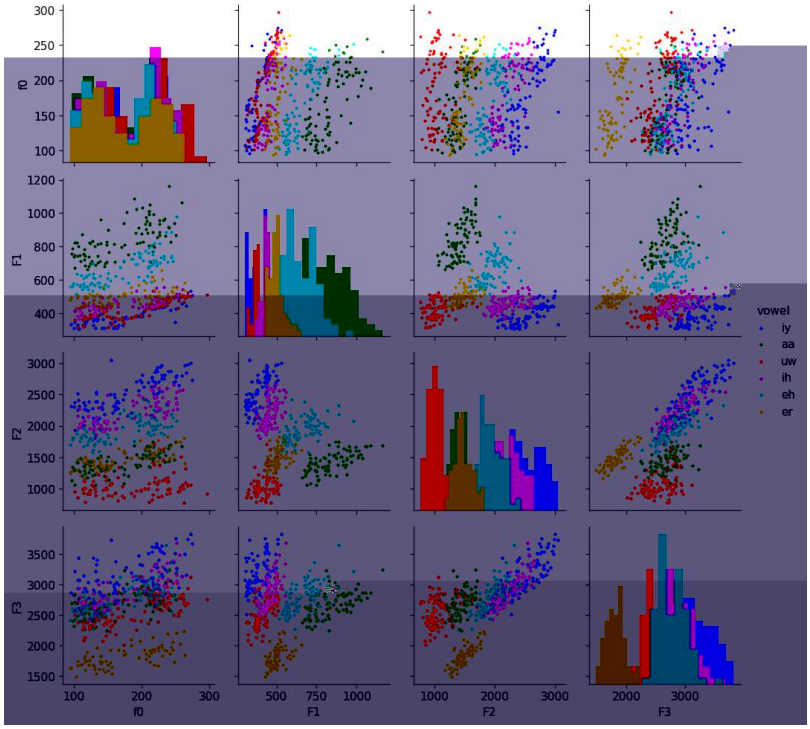
In most of the scatter plots we limited ourselves to the F1-F2 data. However, while F1-F2 may be some of the most relevant features of speech, there is much more: F3, f0, ...

The 2D scatter plot has its limits if we want to see how all these features can work together. A simple extension of the scatter plot is the grid plot which combines a multitude of 2D scatter plots. It doesn't show the true high-dimensional distribution but a number of 2D sub views which already can tell quite a bit more.

It is quite obvious that extra information may be obtained from F3 and f0.

- f0 has a rather clean bimodal distribution, which correlates well to gender
- F3 is in a few cases very complementary, e.g. for /er/, which is not well distinguished with F1,F2 alone

```
<Figure size 1000x1000 with 0 Axes>
```



Out[16]: `<seaborn.axisgrid.FacetGrid at 0x19d780e09a0>`