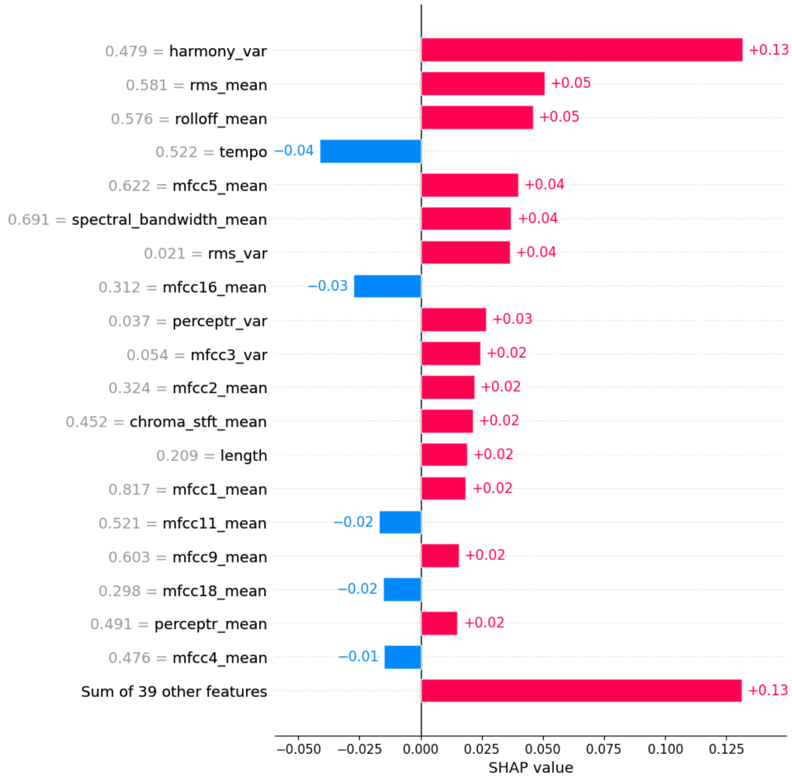
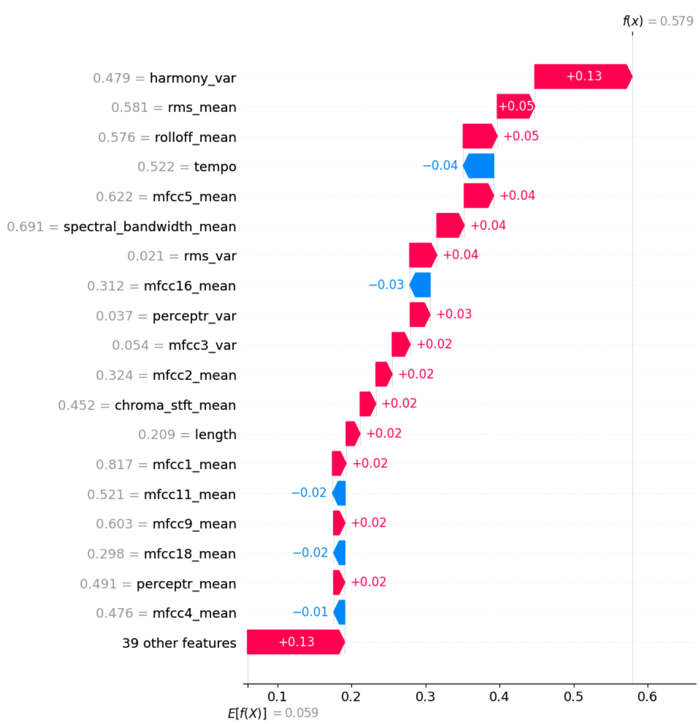
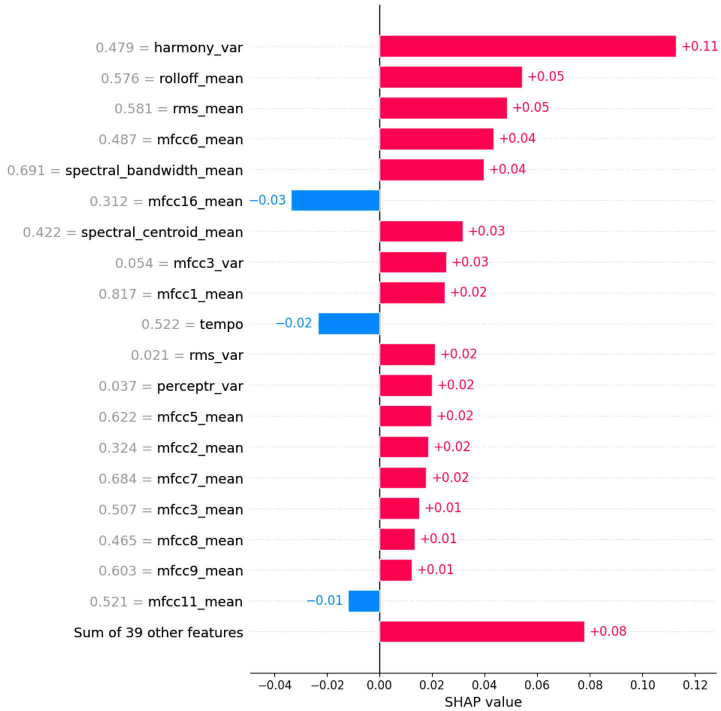
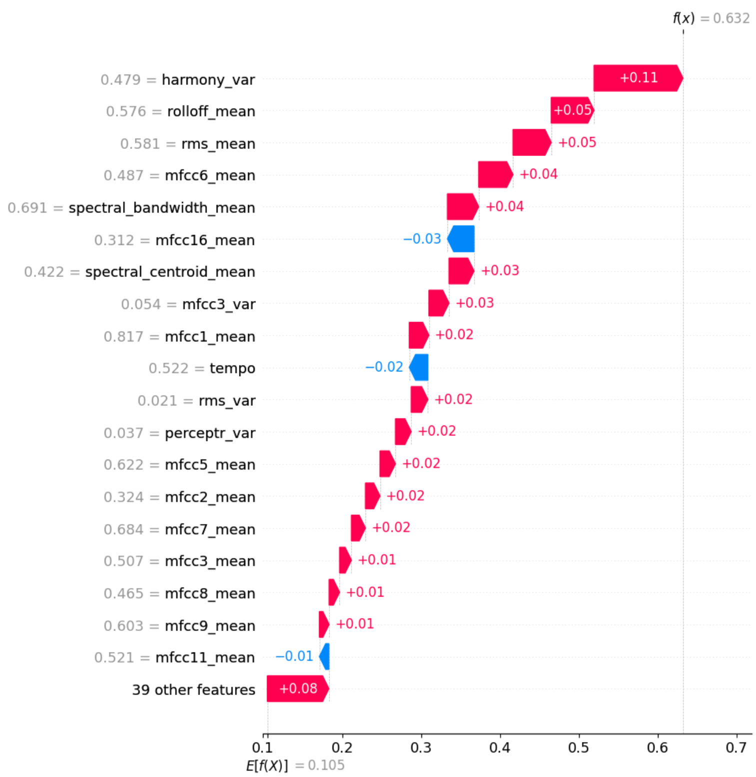
Results
30 seconds





3 seconds





Red Bars (Positive SHAP Values)
Push prediction higher
Increase model output
Example from your plot:

harmony_var +0.13 → strongest positive contributor
rolloff_mean +0.08
spectral_bandwidth_mean +0.05
These features are increasing the model's prediction.

Blue Bars (Negative SHAP Values)
Push prediction lower
Decrease model output
Example:
tempo -0.05
mfcc16_mean -0.05
mfcc18_mean -0.02
These features reduce the prediction.

Tempo/BPM | sum(|SHAP|) @30s = 0.0412 vs @3s = 0.0236 due to noisiness from the clips, hence a lower absolute value for 3 seconds

At 30 seconds, KNN recognizes Country as **rhythmic continuity**.
At 3 seconds, KNN recognizes Country as **spectral resemblance**.

At **30s**:
Genre is interpreted as **temporal form**.
The model recognizes:
- Groove
- Rhythmic pacing
- Harmonic continuity

At **3s**:
Genre is interpreted as **spectral snapshot**.
The model recognizes:
- Frequency distribution
- Attack characteristics
- Timbral spikes

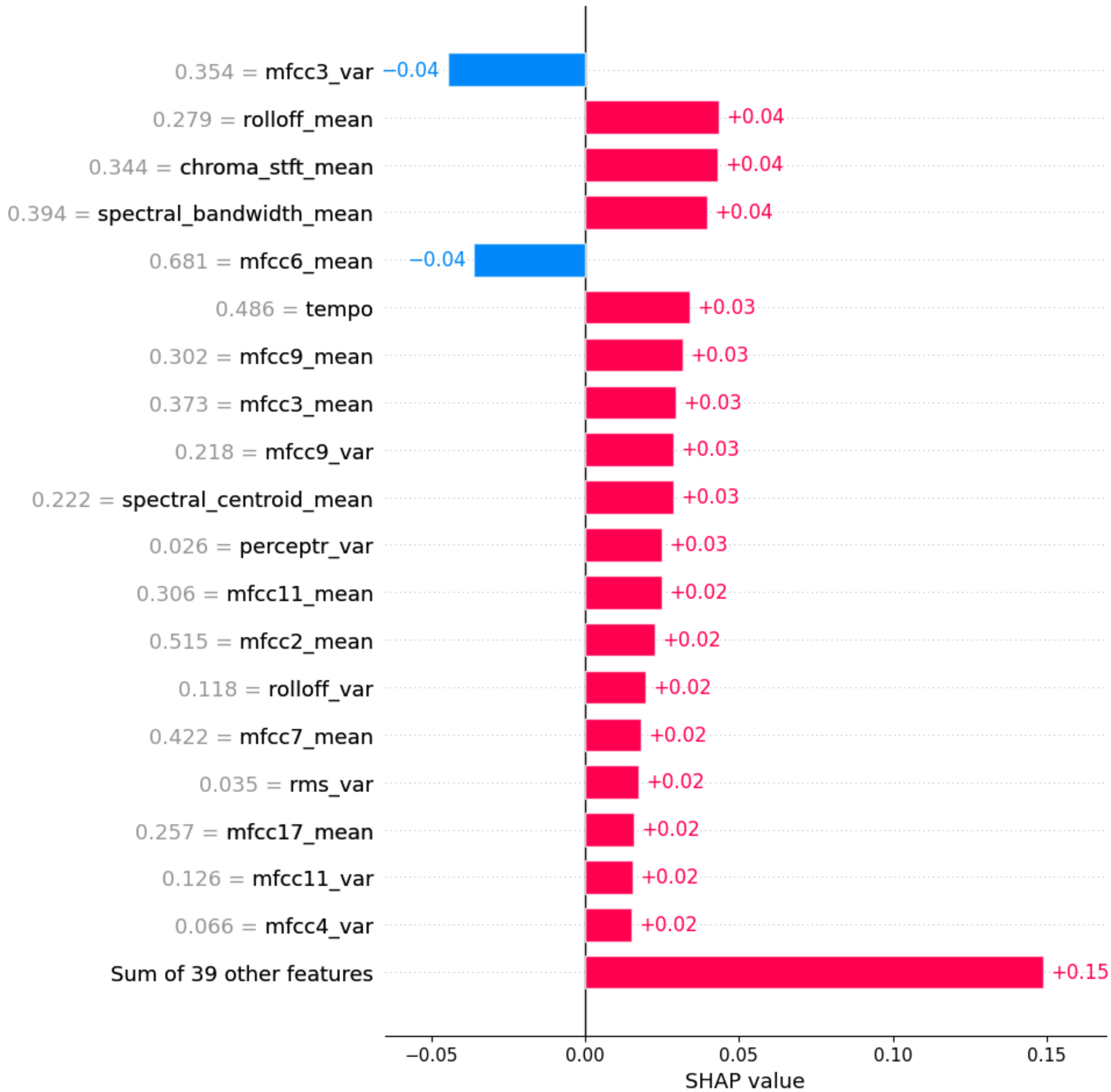
30s Summary Plot
- Tempo high importance
- Tight SHAP distribution
- Harmonic averages matter

3s Summary Plot
- Tempo low importance
- Wide SHAP variance
- MFCC variance and spectral spikes dominate

Using 30 secs model misclassified file jazz.00008.1.wav

	rank	distance	neighbor_label	neighbor_filename
0	1	0.381925	jazz	jazz.00008.2.wav
1	2	0.381925	jazz	jazz.00008.2.wav
2	3	0.449899	country	country.00073.5.wav
3	4	0.449899	country	country.00073.5.wav
4	5	0.453945	country	country.00083.6.wav
5	6	0.453945	country	country.00083.6.wav
6	7	0.483940	country	country.00074.7.wav
7	8	0.483940	country	country.00074.7.wav
8	9	0.487872	country	country.00074.5.wav
9	10	0.487872	country	country.00074.5.wav
10	11	0.491250	blues	blues.00002.5.wav
11	12	0.491250	blues	blues.00002.5.wav

Neighbor label counts:
neighbor_label
country 11
disco 4
jazz 2
blues 2
Name: count, dtype: int64
Pred: country | Top-3: [('country', np.float64(0.5789473684210527)), ('disco', np.float64(0.21052631578947367)), ('blues', np.float64(0.10526315789473684))]



tx = gt30 if source=="30s" else gt3

x = ctx["X_test"].loc[idx]

print(f"Using {source} model. Misclassified file:", fname)

Neighbors (to show 'proximity' to Country cluster)

neigh = neighbor_table(ctx, x)

High spectral rolloff = more high-frequency energy

Country music often has bright acoustic guitars.

This Jazz track likely had similar brightness.

Chroma features represent harmonic content.

The harmonic structure resembled Country patterns.

Spectral bandwidth refers to a bright

Mfcc3 and Mfcc6_mean have a negative SHAP value, meaning that they contribute to a predicted value of jazz and not country.

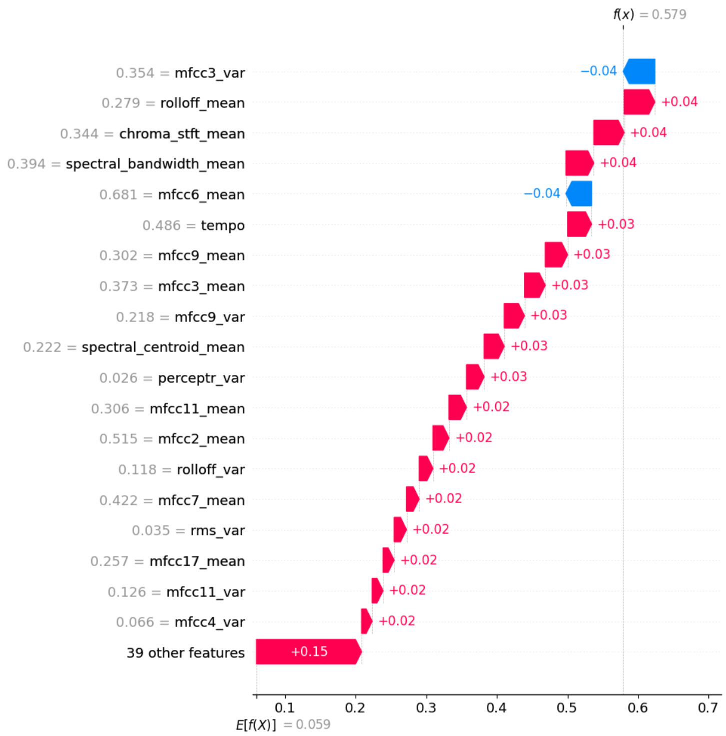
Sum of 39 other features = +0.15 -> Indicates many features collectively push the prediction towards country since the combined effect dominated versus a single feature prediction

Many small features collectively pushed the prediction strongly toward Country.

Even though no single feature was huge, their combined effect dominated.

This is often why misclassifications happen.

The Jazz track was misclassified because its spectral brightness and harmonic features (e.g., high rolloff, bandwidth, and chroma) looked more like Country than Jazz, so the combined feature signals pushed the model toward Country.



Centroid Sum [SHAP] 0.036184210526315 88
Chroma Sum [SHAP] 0.05269736842105260 4

Since Chroma sum(|SHAP|) = 0.0527 is close to 1.5 times the Spectral Centroid sum(|SHAP|) = 0.0362, the misclassification was driven mainly by harmonic (chroma) features, not brightness (centroid).

So the model confused this Jazz track with Country primarily because its harmonic structure looked more like Country.