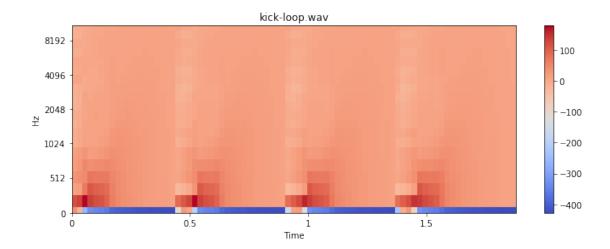
neural_network

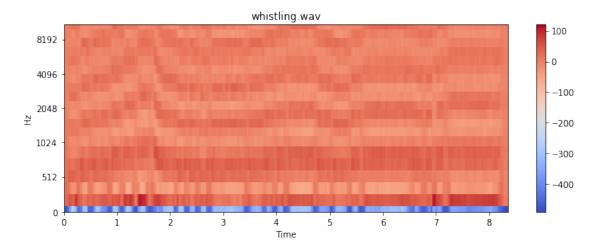
September 27, 2019

```
In [1]: import librosa
        import librosa.feature
        import librosa.display
        import glob
        import numpy as np
        import matplotlib.pyplot as plt
        from keras.models import Sequential
        from keras.layers import Dense, Activation
        from keras.utils.np_utils import to_categorical
        #Using TensoFlow backend.
Using TensorFlow backend.
d:\dev\python\python36\lib\site-packages\tensorflow\python\framework\dtypes.py:523: FutureWarn
  _np_qint8 = np.dtype([("qint8", np.int8, 1)])
d:\dev\python\python36\lib\site-packages\tensorflow\python\framework\dtypes.py:524: FutureWarn
  _np_quint8 = np.dtype([("quint8", np.uint8, 1)])
d:\dev\python\python36\lib\site-packages\tensorflow\python\framework\dtypes.py:525: FutureWarn
  _np_qint16 = np.dtype([("qint16", np.int16, 1)])
d:\dev\python\python36\lib\site-packages\tensorflow\python\framework\dtypes.py:526: FutureWarn
  _np_quint16 = np.dtype([("quint16", np.uint16, 1)])
d:\dev\python\python36\lib\site-packages\tensorflow\python\framework\dtypes.py:527: FutureWarn
  _np_qint32 = np.dtype([("qint32", np.int32, 1)])
d:\dev\python\python36\lib\site-packages\tensorflow\python\framework\dtypes.py:532: FutureWarn
 np_resource = np.dtype([("resource", np.ubyte, 1)])
In [5]: def display_mfcc(song):
           y, _ = librosa.load(song)
           mfcc = librosa.feature.mfcc(y)
           plt.figure(figsize=(10,4))
           librosa.display.specshow(mfcc, x_axis='time', y_axis='mel')
           plt.colorbar()
           plt.title(song)
           plt.tight_layout()
           plt.show()
```

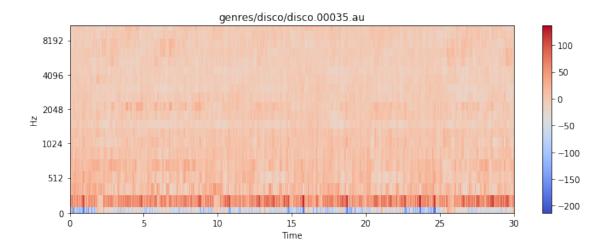
In [9]: display_mfcc('kick-loop.wav')



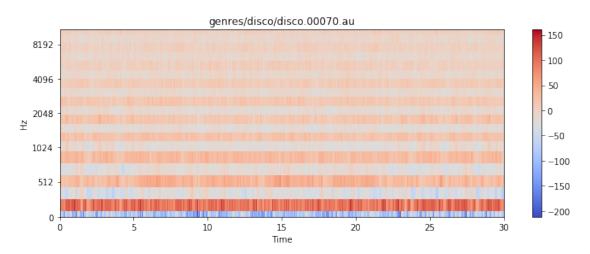
In [12]: display_mfcc('whistling.wav')



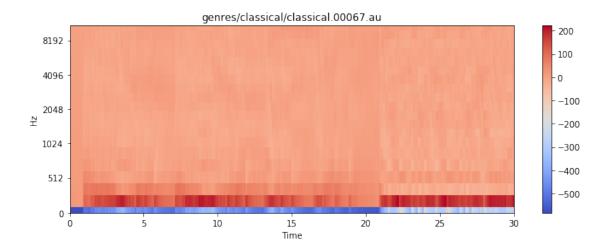
In [13]: display_mfcc('genres/disco/disco.00035.au')



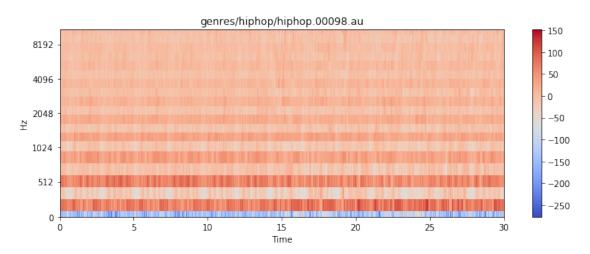
In [14]: display_mfcc('genres/disco/disco.00070.au')



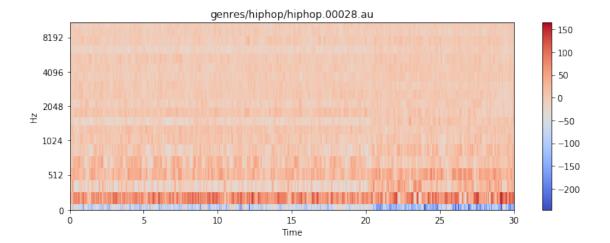
In [15]: display_mfcc('genres/classical/classical.00067.au')



In [16]: display_mfcc('genres/hiphop/hiphop.00098.au')



In [17]: display_mfcc('genres/hiphop/hiphop.00028.au')



```
In [2]: def extract_features_song(f):
            y, _=librosa.load(f)
            # get Mel-frequency cepstral coefficients
           mfcc = librosa.feature.mfcc(y)
            # normalize values between -1, 1 (divide by max)
           mfcc /= np.amax(np.absolute(mfcc))
            return np.ndarray.flatten(mfcc)[:25000]
In [3]: def generate_features_and_labels():
            all_features = []
            all_labels = []
            genres = ['blues', 'classical', 'country', 'hiphop', 'jazz', 'metal', 'pop', 'regg
            for genre in genres:
                sound_files = glob.glob('genres/'+genre+'/*.au')
                print('Processing %d songs in %s genre...' % (len(sound_files), genre))
                for f in sound files:
                    features = extract_features_song(f)
                    all_features.append(features)
                    all_labels.append(genre)
            # convert labels to one-hot encoding
            label_uniq_ids, label_row_ids = np.unique(all_labels, return_inverse=True)
            label_row_ids = label_row_ids.astype(np.int32, copy=False)
            onehot_labels = to_categorical(label_row_ids, len(label_uniq_ids))
            return np.stack(all_features), onehot_labels
In [4]: features, labels = generate_features_and_labels()
Processing 100 songs in blues genre...
Processing 100 songs in classical genre...
```

```
Processing 100 songs in country genre...
Processing 100 songs in hiphop genre...
Processing 100 songs in jazz genre...
Processing 100 songs in metal genre...
Processing 100 songs in pop genre...
Processing 100 songs in reggae genre...
Processing 100 songs in rock genre...
In [5]: print(np.shape(features))
        print(np.shape(labels))
        training_split = 0.8
        # last column has genres, turn it in to unique ids
        alldata = np.column_stack((features, labels))
        np.random.shuffle(alldata)
        splitidx = int(len(alldata) * training_split)
        train, test = alldata[:splitidx,:], alldata[splitidx:, :]
        print(np.shape(train))
        print(np.shape(test))
        train input = train[:, :-10]
        train_labels = train[:,-10:]
        test_input = test[:, :-10]
        test_labels = test[:, -10:]
        print(np.shape(train_input))
        print(np.shape(train_labels))
(900, 25000)
(900, 9)
(720, 25009)
(180, 25009)
(720, 24999)
(720, 10)
In [6]: model = Sequential([
            Dense(100, input_dim=np.shape(train_input)[1]),
            Activation('relu'),
            Dense(10),
            Activation('softmax'),
        ])
```

```
model.compile(optimizer='adam',
        loss='categorical_crossentropy',
        metrics=['accuracy'])
   print(model.summary())
   model.fit(train_input, train_labels, validation_split=0.2, epochs=10, batch_size=32)
   loss, acc = model.evaluate(test_input, test_labels, batch_size=32)
   print("Done!")
   print("Loss: %.4f, accuracy: %.4f" % (loss, acc))
Layer (type)
       Output Shape
                 Param #
______
         (None, 100)
dense_1 (Dense)
                     2500000
activation_1 (Activation) (None, 100)
_____
dense_2 (Dense)
          (None, 10)
                    1010
______
activation 2 (Activation) (None, 10)
Total params: 2,501,010
Trainable params: 2,501,010
Non-trainable params: 0
-----
Train on 576 samples, validate on 144 samples
Epoch 1/10
Epoch 2/10
Epoch 3/10
Epoch 4/10
Epoch 5/10
Epoch 6/10
Epoch 7/10
Epoch 8/10
Epoch 9/10
```

```
Epoch 10/10
180/180 [=========== ] - Os 189us/step
Loss: 1.2868, accuracy: 0.6389
In [7]: import pandas as pd
       from keras.preprocessing.text import Tokenizer
       import numpy as np
       from keras.models import Sequential
       from keras.layers import Dense, Dropout, Activation
       from keras.utils import np_utils
       from sklearn.model_selection import StratifiedKFold
In [8]: d=pd.concat([pd.read_csv("youtube-spam/Youtube01-Psy.csv"),
                    pd.read_csv("youtube-spam/Youtube02-KatyPerry.csv"),
                    pd.read_csv("youtube-spam/Youtube03-LMFAO.csv"),
                    pd.read_csv("youtube-spam/Youtube04-Eminem.csv"),
                    pd.read_csv("youtube-spam/Youtube05-Shakira.csv")])
       d = d.sample(frac=1)
In [9]: kfold = StratifiedKFold(n_splits=5)
       splits = kfold.split(d, d['CLASS'])
In [10]: for train, test in splits:
            print("Split")
            print(test)
Split
  0
      1
          2
              3
                  4
                      5
                          6
                             7
                                 8
                                     9
                                        10
                                           11
                                                12
                                                   13
                                                       14
                                                           15
                                                               16
                                                                   17
  18
     19
         20 21
                 22
                     23
                        24
                             25
                                26
                                    27
                                        28
                                            29
                                                30
                                                    31
                                                        32
                                                           33
                                                               34
                                                                   35
  36
     37
         38
             39
                 40
                     41
                        42
                             43
                                44
                                    45
                                        46
                                            47
                                                48
                                                    49
                                                        50
                                                           51
                                                               52
                                                                   53
  54
                                                                   71
     55
         56
            57
                 58
                     59
                        60
                             61
                                62
                                    63
                                        64
                                            65
                                                66
                                                    67
                                                        68
                                                           69
                                                               70
  72
     73
         74
            75
                 76
                     77
                         78
                            79
                                80
                                        82
                                            83
                                                84
                                                    85
                                                           87
                                                                   89
                                    81
                                                       86
                                                               88
     91
         92
             93
                 94
                     95
                         96
                            97
                                98
                                    99 100 101 102 103 104 105 106 107
 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125
 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143
 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161
 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179
 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197
 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215
 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233
 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251
 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269
 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287
 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305
 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323
```

```
1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076
 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090
 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104
 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118
 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132
 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146
 1147 1149 1150 1152 1153 1154 1156 1157 1158 1159 1160 1161 1162 1164
 1166 1168 1169 1171 1172 1173 1174 1175 1178 1180 1183 1185 1188
Split
[1148 1151 1155 1163 1165 1167 1170 1176 1177 1179 1181 1182 1184 1186
 1187 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201
 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215
 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229
 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243
 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257
 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271
 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285
 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299
 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313
 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327
 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341
 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355
 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369
 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383
 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397
 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411
 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425
 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439
 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453
 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467
 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481
 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495
 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509
 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523
 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537
 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1555
 1557 1559 1560 1561 1563 1565 1566 1567 1570 1571 1572 1573 1575]
Split
[1551 1552 1553 1554 1556 1558 1562 1564 1568 1569 1574 1576 1577 1578
 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592
 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606
 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620
 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634
 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648
 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662
 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676
 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690
 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704
```

```
1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718
 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732
 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746
 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760
 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774
 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788
 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802
 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816
 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830
 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844
 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858
 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872
 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886
 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900
 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914
 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928
 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942
 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955]
In [23]: def train_and_test(train_idx, test_idx):
             train_content = d['CONTENT'].iloc[train_idx]
             test_content = d['CONTENT'].iloc[test_idx]
             tokenizer = Tokenizer(num words=2000)
             # Learn the training words (not the testing words!)
             tokenizer.fit_on_texts(train_content)
             # options for made : binary, freq, tfidf
             d_train_inputs = tokenizer.texts_to_matrix(train_content, mode='tfidf')
             d_test_inputs = tokenizer.texts_to_matrix(test_content, mode='tfidf')
             # divide tfidf by max
             d_train_inputs = d_train_inputs/np.amax(np.absolute(d_train_inputs))
             d_test_inputs = d_test_inputs/np.amax(np.absolute(d_test_inputs))
             # substract mean, to get values between -1 and 1
             d_train_inputs = d_train_inputs - np.mean(d_train_inputs)
             d_test_inputs = d_test_inputs - np.mean(d_test_inputs)
             # one-hot encoding of outputs
             d_train_outputs = np_utils.to_categorical(d['CLASS'].iloc[train_idx])
             d_test_outputs = np_utils.to_categorical(d['CLASS'].iloc[test_idx])
            model = Sequential()
             model.add(Dense(512, input_shape=(2000,)))
             model.add(Activation('relu'))
```

```
model.add(Dropout(0.5))
     model.add(Dense(2))
     model.add(Activation('softmax'))
     model.compile(loss='categorical crossentropy', optimizer='adamax',
          metrics=['accuracy'])
     model.fit(d_train_inputs, d_train_outputs, epochs=10, batch_size=16)
     scores = model.evaluate(d_test_inputs, d_test_outputs)
     print("%s: %.2f%%" % (model.metrics_names[1], scores[1]*100))
     return scores
In [24]: kfold = StratifiedKFold(n_splits=5)
   splits = kfold.split(d, d['CLASS'])
   cvscores = []
   for train_idx, test_idx in splits:
     scores = train_and_test(train_idx, test_idx)
     cvscores.append(scores[1]*100)
Epoch 1/10
Epoch 2/10
Epoch 3/10
Epoch 4/10
Epoch 5/10
Epoch 6/10
Epoch 7/10
Epoch 8/10
Epoch 9/10
Epoch 10/10
392/392 [========== ] - Os 270us/step
acc: 96.17%
Epoch 1/10
Epoch 2/10
Epoch 3/10
Epoch 4/10
```

```
Epoch 5/10
Epoch 6/10
Epoch 7/10
Epoch 8/10
Epoch 9/10
Epoch 10/10
391/391 [========= ] - 0s 314us/step
acc: 95.40%
Epoch 1/10
Epoch 2/10
Epoch 3/10
Epoch 4/10
Epoch 5/10
Epoch 6/10
Epoch 7/10
Epoch 8/10
Epoch 9/10
Epoch 10/10
391/391 [=========== ] - 0s 338us/step
acc: 94.37%
Epoch 1/10
Epoch 2/10
Epoch 3/10
Epoch 4/10
Epoch 5/10
Epoch 6/10
```

```
Epoch 7/10
Epoch 8/10
Epoch 9/10
Epoch 10/10
391/391 [========= ] - Os 371us/step
acc: 94.63%
Epoch 1/10
Epoch 2/10
Epoch 3/10
Epoch 4/10
Epoch 5/10
Epoch 6/10
Epoch 7/10
Epoch 8/10
Epoch 9/10
Epoch 10/10
391/391 [========== ] - Os 384us/step
acc: 95.40%
```

In []: