

## MFCC feature Extraction

- The MFCC features were extracted using the code provided by the professor.
- The final values are stored in a tensor of dimensions 39x123xN
- New test audios were recorded as instructed and the euclidian distance was calculated.
- The accuracy rate of both of our test audios were 89.7%.
- The accuracy rate of five other individuals is 58.3%.

```
clc
clear all
files = ["car_N1.m4a", "car_N2.m4a", "car_N3.m4a", "car_N4.m4a", "car_N5.m4a", "car_S1.m4a", "car_S2.m4a"];
for Speechindex = 1:length(files)
    speech = audioread(files(Speechindex));
    A(Speechindex) = length(speech);
    speech = speech(1:20000,1);
    fs = 16000;
    Tw = 25;
    Ts = 10;
    alpha = 0.97;
    R = [ 50 3000 ];
    M = 20;
    C = 12;
    L = 22;
    hamming = @(N)(0.54-0.46*cos(2*pi*[0:N-1].'/(N-1)));
    [ MFCCs, FBEs, frames,eframes ] = mfcc(speech, fs, Tw, Ts, alpha, hamming, R, M, C, L );
    for i = 1:size(MFCCs,2)
        if and(i > 1,i<size(MFCCs,2))
            MFCCd(:,i) = (MFCCs(:,i+1) - MFCCs(:,i-1))/2;
            eframesd(i) = (eframes(i+1) - eframes(i-1))/2;
        else
            MFCCd(:,i) = MFCCs(:,i);
            eframesd(i) = eframes(i);
        end
    end

    end

    for j = 1:size(MFCCs,2)
        if and(j > 1,j<size(MFCCs,2))
            MFCCdd(:,j) = (MFCCd(:,j+1) - MFCCd(:,j-1))/2;
            eframesdd(i) = (eframesd(j+1) - eframesd(j-1))/2;
        else
            MFCCdd(:,j) = MFCCd(:,j);
            eframesdd(j) = eframesd(j);
        end
    end
    MFCCfeatures_x(1:39,1:123,Speechindex) = [MFCCs(1:12,1:123);MFCCd(1:12,1:123);MFCCdd(1:12,1:123);FBEs(1:12,1:123);frames(1:12,1:123);eframes(1:12,1:123)];
end
```

```
test_files = ["enter files here"];
```

```

for testindex = 1:length(test_files)
    speech_test = audioread(files(testindex));
    speech_test = speech_test(1:20000,1);
    fs1 = 16000;
    Tw1 = 25;
    Ts1 = 10;
    alpha1 = 0.97;
    R1 = [ 50 3000 ];
    M1 = 20;
    C1 = 12;
    L1 = 22;
    hamming1 = @(N)(0.54-0.46*cos(2*pi*[0:N-1].'/(N-1)));
    [ MFCCs1, FBES1, frames1,eframes1 ] = mfcc(speech_test, fs1, Tw1, Ts1, alpha1, hamming1, R1, M1, C1, L1);
    for i = 1:size(MFCCs1,2)
        if and(i > 1,i<size(MFCCs1,2))
            MFCCd1(:,i) = (MFCCs1(:,i+1) - MFCCs1(:,i-1))/2;
            eframesd1(i) = (eframes1(i+1) - eframes1(i-1))/2;
        else
            MFCCd1(:,i) = MFCCs1(:,i);
            eframesd1(i) = eframes1(i);
        end

    end

    for j = 1:size(MFCCs1,2)
        if and(j > 1,j<size(MFCCs1,2))
            MFCCdd1(:,j) = (MFCCd1(:,j+1) - MFCCd1(:,j-1))/2;
            eframesdd1(i) = (eframesd1(j+1) - eframesd1(j-1))/2;
        else
            MFCCdd1(:,j) = MFCCd1(:,j);
            eframesdd1(j) = eframesd1(j);
        end

    end

    MFCCfeatures_x1(1:39,1:123,1) = [MFCCs1(1:12,1:123);MFCCd1(1:12,1:123);MFCCdd1(1:12,1:123)];

    for k = 1:50
        Xcheck = MFCCfeatures_x1(1:39,1:123,1);
        Xcheck(isnan(Xcheck))=0;
        Euc_dist(k) = norm(MFCCfeatures_x(1:39,1:123,k)-(Xcheck).^2 ) ;
    end
    [minD I] = min(Euc_dist);
    X = sprintf('the test word is %s and %s is the spoken word.',test_files(testindex),files(I));
    disp(X)
end

```

```

the test word is testaudiochair.m4a and chair_S5.m4a is the spoken word.
the test word is testaudiocar.m4a and desk_S1.m4a is the spoken word.
the test word is testaudiodesk.m4a and desk_S1.m4a is the spoken word.
the test word is chair1 and desk_S1.m4a is the spoken word.
the test word is chair2 and phone_S1.m4a is the spoken word.
the test word is chair3 and car_S1.m4a is the spoken word.
the test word is chair4 and car_S5.m4a is the spoken word.
the test word is chair5 and car_S5.m4a is the spoken word.

```