



**AYRIŞIK SÖZCÜK TABANLI TÜRKÇE KONUŞMACI TANIMA SİSTEMİ
GELİŞTİRME VE ANAHTAR KELİME SEÇİMİNİN KONUŞMACI
TANIMA PERFORMANSINA ETKİSİNİN İNCELENMESİ**

Zekeriya ŞENTÜRK

**YÜKSEK LİSANS
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

MAYIS 2015

Zekeriya ŞENTÜRK tarafından hazırlanan “AYRIŞIK SÖZCÜK TABANLI TÜRKÇE KONUŞMACI TANIMA SİSTEMİ GELİŞTİRME VE ANAHTAR KELİME SEÇİMİNİN KONUŞMACI TANIMA PERFORMANSINA ETKİSİNİN İNCELENMESİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Elektrik Elektronik Mühendisliği Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Doç. Dr. Özgül SALOR DURNA

Elektrik-Elektronik Mühendisliği Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

İkinci Danışman: Prof. Dr. Önder Haluk TEKBAŞ

Savunma Bilimleri Enstitüsü Müdürlüğü, Kara Harp Okulu

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

Başkan : Prof. Dr. Müzeyyen SARITAŞ

Elektrik-Elektronik Mühendisliği Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

Üye : Doç. Dr. M.Timur AYDEMİR

Elektrik-Elektronik Mühendisliği Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

Üye : Doç. Dr. Nursel AKÇAM

Elektrik-Elektronik Mühendisliği Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

Tez Savunma Tarihi: 27/05/2015

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....
Prof. Dr. Şeref SAĞIROĞLU
Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Zekeriya ŞENTÜRK

27/05/2015

**AYRIŞIK SÖZCÜK TABANLI TÜRKÇE KONUŞMACI TANIMA SİSTEMİ
GELİŞTİRME VE ANAHTAR KELİME SEÇİMİNİN KONUŞMACI TANIMA
BAŞARIMINA ETKİSİNİN İNCELENMESİ**
(Yüksek Lisans Tezi)

Zekeriya ŞENTÜRK

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Mayıs 2015

ÖZET

Bu tez çalışması kapsamında, ayrışık sözcük kullanımıyla bir konuşmacı tanıma sistemi geliştirilmiştir. Geliştirilen sistemde öznitelik olarak Mel Frekans Cepstrum Katsayıları (Mel Frequency Cepstrum Coefficient - MFCC) ve kodlanmış çizgisel frekans spektrumu (Line Spectral Frequency - LSF) kullanılmıştır. Çalışma kapsamında, geliştirilen konuşmacı tanıma sisteminin başarımı, seçilen özniteliklere ve kullanılan anahtar kelimelere göre değerlendirilmiştir. Çalışmanın yapılabilmesi için, 12 erkek ve 4 kadın konuşmacının 48 farklı kelimeyi 6'şar kez tekrarladığı bir veri tabanı oluşturulmuştur. Kayıtlar, ses yalıtımı olan gürültüsüz bir kayıt stüdyosunda; 16 kHz örnekleme frekansı ve Logitech marka kablosuz bir mikrofon ile alınmıştır. Kayıtlar bir konuşmacı için günün farklı saatleri seçilen üç ya da dört oturumda yapılmış ve kişinin ses değişkenliğinin veri tabanına yansıtılması sağlanmıştır. Tezin ilk bölümünde, konuşmacı veri tabanı oluşturulduktan sonra, ikinci bölümde öznitelik olarak MFCC'ler kullanılarak konuşmacı tanıma sistemi oluşturulmuş ve sistemin başarımı ölçülmüştür. Üçüncü bölümde ise, MELP ile kodlanmış ve kod çözülerek tekrar oluşturulmuş konuşma sinyali sisteme giriş olarak uygulanarak, geliştirilen sistemin konuşmacı tanıma başarımının kodlamaya karşı gürbüzlüğü değerlendirilmiştir. Dördüncü olarak, kodlanmış ve kod çözülmüş sözcüklerin konuşmacı tanıma sisteminde yol açtığı başarımların düşüşünü azaltmak üzere, öznitelik olarak MELP konuşma kodlama algoritmasının kod sözcüklerini kullanan bir konuşmacı tanıma sistemi geliştirilmiştir. Yalnızca MFCC'ler öznitelik olarak kullanıldığı zaman %96 olarak elde edilen başarımın, kodlanmış ve kod çözülmüş sözcükler sisteme giriş olarak verildiğinde başarımın %60'a düştüğü görülmüştür. Kodlanmaya karşı dayanıklı bir sistem geliştirmek için oluşturulan ve MELP kod sözcüklerinin kullanan sistemin konuşmacı tanıma başarımının ise %71'e yükseldiği gözlenmiştir. Son bölümde ise, geliştirilen temel sistemin seçilen anahtar sözcüğe göre başarımı ölçülmüştür. Yapılan testler sonucunda, sesletimi kolay, yabancı kökenli olmayan sözcüklerde konuşmacı tanıma başarımının, diğer sözcüklere göre daha yüksek olduğu gözlenmiştir.

Bilim Kodu : 905.1.067
Anahtar Kelimeler : Konuşmacı tanıma, LSF, MELP, MFCC
Sayfa Adedi : 61
Danışman : Doç. Dr. Özgül SALOR DURNA
İkinci Danışman : Prof. Dr. Önder Haluk TEKBAŞ

DEVELOPING AN ISOLATED WORD BASED TURKISH SPEAKER RECOGNITION
SYSTEM AND INVESTIGATION OF THE KEYWORD CHOICE ON THE SPEAKER
RECOGNITION PERFORMANCE

(M. Sc. Thesis)

Zekeriya ŞENTÜRK

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

May 2015

ABSTRACT

In this thesis, an isolated-word-based speaker recognition system is developed. Mel Frequency Cepstrum Coefficients (MFCC) and Line Spectral Frequencies (LSF) are used as the features of the developed speaker recognition system. This system's performance is evaluated for both keyword choice and feature selection. In order to perform this study, a database has been created consisting of 48 words per speaker taken from 12 men and 4 women and each word is recorded 6 times for each participant. Records are acquired at 16 kHz sampling frequency in a professional recording studio, which is noiseless and has sound insulation, using Logitech wireless microphone. The records acquired 3 or 4 sessions, which were achieved at different times of the day, for each participant to reflect the sound variability of the human vocal tract on the database. In the first part of the thesis, the speaker database has been prepared. In the second section, speaker recognition system has been developed and the system performance has been measured for the system operating with MFCC as the attributes. In the third section, system performance and robustness of the system against coding has been evaluated when MELP coded and decoded speech is provided to the system as the input signal. In the fourth part of the thesis, to increase the robustness of the system against coding, a new speaker recognition system has been proposed, using the MELP coder's codebook as features. When MFCCs are used for speaker recognition as features, the performance of the system has been 96%; however, if coded and decoded speech signal is used as dataset on the same system, the performance of the system has been reduced to 60%. For the speaker recognition system, which uses MELP coder's codebook as features, the performance of the system has been found to be 71%. In the final section, the performance of the baseline system is analyzed with respect to the keyword choice. It has been observed that the words, which have easy pronunciation and are not foreign origin, show higher performance than the others.

Science Code : 905.1.067

Key Words : Speaker Recognition, LSF, MELP, MFCC

Page Number : 61

Supervisor : Assoc. Prof. Dr. Özgül SALOR DURNA

Co-Supervisor : Prof. Dr. Önder Haluk TEKBAŞ

TEŞEKKÜR

Tezimin hazırlık aşamasından son halini verdiğim ana kadar büyük bir özveri ve sabır ile benden desteklerini esirgemeyen saygıdeğer danışman hocam Doç. Dr. Özgül SALOR DURNA hanımefendiye minnetle teşekkür ederim.

Eş danışmanım Sayın Müh.Alb.Prof.Dr.Önder Haluk TEKBAŞ'a yaptığım çalışma boyunca verdiği desteklerden dolayı teşekkürlerimi sunarım.

Tezimde kullandığım ses kayıtlarını oluşturmamda yardımcı olan değerli arkadaşlarıma teşekkür ederim.

Hayatımın her döneminde destekleri, sevgileri ve fedakârlıkları ile yanımda olan aileme ve tezimin her aşamasında bana desteğini esirgememesinin yanında fikirleri ile de yön veren sevgili eşim Elif ŞENTÜRK'e teşekkürü bir borç bilirim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
RESİMLERİN LİSTESİ	xii
SİMGELER VE KISALTMALAR.....	xiii
1. GİRİŞ.....	1
2. KURAMSAL TEMELLER	5
2.1. Konuşma Sinyalinin Oluşumu	5
2.2. Konuşma Sinyalinin Özellikleri	6
2.2.1. Ötümlü ve ötümsüz sesler	6
2.2.2. Perde (Pitch) periyodu	7
2.2.3. Formant frekansları	7
2.3. Konuşma Sinyalinin Modellenmesi	8
2.3.1. Doğrusal öngörücü katsayıları (Linear Prediction Coefficients - LPC) ...	8
2.3.2. Çizgisel spektrum frekansları (Line Spectrum Frequencies - LSF).....	10
2.3.3. Kepstrum katsayıları	11
2.3.4. Mel frekansı kepsrum katsayıları (MFCC).....	12
2.4. Konuşma Sinyalinin Sınıflandırılması	15
2.4.1. Dinamik zaman eğirme (Dynamic Time Warping - DTW).....	15
2.4.2. Vektör Nicemleme (Vector Quantization - VQ).....	16

3. KONUŞMACI VERİTABANININ OLUŞTURULMASI.....	17
4. KONUŞMACI TANIMA SİSTEMİ	21
4.1. MFCC Öznitelikleri Kullanılarak Sınıflandırma Başarımı	22
4.1.1. Test verisine farklı desibel (db) değerlerinde sinyal gürültü oranını (SNR) sağlayacak beyaz Gauss gürültü eklenmesinin sistem performansına etkisi.....	24
4.2. MELP Kodlanıp Kod Açılmış Konuşma Veritabanıyla MFCC Öznitelikleri Kullanarak Sınıflandırma Başarımı.....	25
4.3. MELP Öznitelikleri Kullanılarak Sınıflandırma Başarımı	28
5. MFCC ÖZNİTELİKLERİNE DAYALI KONUŞMACI TANIMADA ANAHTAR KELİME SEÇİMİNİN PERFORMANSA ETKİSİ	35
5.1. Sınıflandırmada %100 Başarı Sağlayan Kelimelerin Fonetik Analizi.....	36
5.2. Sınıflandırmada %100 Başarı Sağlamayan Kelimelerin Fonetik Analizi.....	48
6. SONUÇ VE ÖNERİLER	51
KAYNAKLAR	53
EKLER.....	57
EK-1. 4'ü kadın 12'si erkek 16 konuşmacıyla yapılan test aşamasından sonra tanıma sonuçlarının işlendiği formlar	58
ÖZGEÇMİŞ	61

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Veri tabanına katkı sağlayan konuşmacıların bilgileri	19
Çizelge 3.2. Veri tabanında bulunan 48 farklı kelime ve veri tabanı indisleri.....	19

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. İnsanın konuşma üretim sisteminin çalışması	5
Şekil 2.2. ‘Rana sarı bir zambak giyecek mi?’ zaman bölgesi genlik ve Laringograf sinyali gösterimi.....	6
Şekil 2.3. Konuşma sinyali oluşumunun sayısal süzgeç benzetimi	8
Şekil 2.4. Konuşma sinyali LPC sentezi modeli	9
Şekil 2.5. Çizgisel frekans çiftinin çıkarımı.....	10
Şekil 2.6. Konuşma sinyalinin kısa zamanlı kepsral analizi.....	12
Şekil 2.7. Mel – frekans dönüşüm eğrisi.....	12
Şekil 2.8. MFCC özniteliklerinin çıkarımı	13
Şekil 2.9. MFCC öznitelik çıkarımı için kullanılan süzgeç takımı.....	14
Şekil 2.10. Aynı niteliklere sahip farklı uzunluktaki iki verinin eşleştirilmesi	15
Şekil 4.1. Konuşmacı tanıma sistemi.....	21
Şekil 4.2. Veri tabanındaki 4 kişiye ait “dümdüz” sözcüğünün MFCC yüzeyi.....	22
Şekil 4.3. MFCC öznitelikleri kullanılarak yapılan sınıflandırmada kişilerin yüzde başarımları.....	23
Şekil 4.4. MFCC öznitelikleri kullanılarak yapılan sınıflandırmada kelimelerin yüzde başarımları.....	23
Şekil 4.5. Sistemin testi için kullanılacak sinyallerin farklı SNR değerlerinde sistem başarımları.....	24
Şekil 4.6. MELP kod açıcı blok diyagramı.....	26
Şekil 4.7. MELP kodlanıp kod açılmış veri tabanının oluşturulması	26
Şekil 4.8. MELP kodlanıp kod açılmış konuşma veritabanıyla MFCC öznitelikleri kullanarak yapılan sınıflandırmada kişiler için yüzde başarımları	27
Şekil 4.9. MELP kodlanıp kod açılmış konuşma veritabanıyla MFCC öznitelikleri kullanarak yapılan sınıflandırmada kelimeler için yüzde başarımları	27

Şekil	Sayfa
Şekil 4.10. Konuşmacı tanıma sisteminde MELP kodlamanın öznitelik çıkarımında kullanılması.....	29
Şekil 4.11. MELP kodlamanın öznitelik çıkarımında kullanılması	29
Şekil 4.12. Çok katmanlı MELP kod kitabından alınmış örnek vektörler	30
Şekil 4.13. Çok katmanlı MELP kod kitabından nicemlenmiş LSF vektörünün çıkarımı	30
Şekil 4.14. Çok katmanlı kod kitabının katmanlarındaki değerlerin süzgeç cevabına katkısı	31
Şekil 4.15. Öznitelik olarak MELP kodlamadan elde edilen nicemlenmiş LSF değerlerinin kullanılmasıyla yapılan sınıflandırmada kişiler için yüzde başarımlar	32
Şekil 4.16. Öznitelik olarak MELP kodlamadan elde edilen nicemlenmiş LSF değerlerinin kullanılmasıyla yapılan sınıflandırmada kelimeler için yüzde başarımlar	33
Şekil 5.1. “Bıçak” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi	36
Şekil 5.2. Veri tabanındaki 4 kişiye ait “Bıçak” sözcüğünün MFCC yüzeyi	37
Şekil 5.3. “Bozuk” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi ...	37
Şekil 5.4. “Dilek” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi.....	38
Şekil 5.5. “Dümdüz” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi	39
Şekil 5.6. “Gıda” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi.....	40
Şekil 5.7. “Güneş” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi ...	41
Şekil 5.8. “Pırasa” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi	42
Şekil 5.9. “Polis” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi.....	43
Şekil 5.10. “Pörsük” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi	44
Şekil 5.11. “Püre” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi	45
Şekil 5.12. “Tırnak” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi .	46
Şekil 5.13. “Toprak” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi	47
Şekil 5.14. “Tuzak” sözcüğünün sırasıyla zaman bölgesi ve MFCC yüzeyi gösterimi..	48

RESİMLERİN LİSTESİ

Resim	Sayfa
Resim 3.1. Kayıtların alındığı profesyonel stüdyo	17
Resim 3.2. Stüdyo duvarlarındaki yalıtım malzemesi.....	18
Resim 3.3. Kayıt alımı sırasında baş üstü mikrofonun konumu	18

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

dB

Decibel

Hz

Hertz

kHz

Kilo Hertz

ms

Mili saniye

Kısaltmalar

Açıklamalar

DCT

Discrete Cosine Transform

DFT

Discrete Fourier Transform

DTW

Dynamic Time Warping

FFT

Fast Fourier Transform

FIR

Finite Impulse Response

LPC

Linear Prediction Coefficients

LSF

Line Spectral Frequency

MELP

Mixed Excitation Linear Prediction

MFCC

Mel Frequency Cepstral Coefficients

SNR

Signal to Noise Ratio

1. GİRİŞ

Parmak izi, yüz, iris, el geometrisi, imza gibi insana ait biyometrik özellikler bilim insanlarının dikkatlerini çekmiş ve üzerlerinde birçok araştırma yapılmıştır. Konuşma sinyali de insanın biyometrik özelliklerinden biridir ve konuşma sinyali üzerinde yıllardır süregelen araştırmalar sayısal bilgisayarların gelişmesi ile birlikte daha da hız kazanmıştır. Konuşma işleme, haberleşme, güvenlik, istihbarat, dilbilim, biyomedikal, kontrol, psikoloji gibi birçok alanın ortak paydası olarak işlev görmektedir. Haberleşme alanında sesin kodlanması ve kod çözülmesi, güvenlik alanında konuşmacı tanıma sistemlerinin yapılması, istihbarat alanında sözcük yakalama uygulamaları, dilbilim alanında ses üretim mekanizmasının incelenmesi, biyomedikal alanında konuşma iyileştirme uygulamaları, kontrol alanında sesli kontrol sistemleri, psikoloji alanında konuşma sinyali ile ruhsal durum tespiti bu alanlara örnek olarak verilebilir.

Konuşma işleme kendi içerisinde birçok alt başlığa ayrılırken, özellikle güvenlik uygulamalarında konuşmacı tanıma sistemleri hala üzerinde çalışılan bir konu olarak yer almaktadır. Konuşmacı tanıma, önceden alınmış konuşma verilerinin özniteliklerinin modellenmesi ile oluşturulan kimlik verilerinin, yeni gelecek bir konuşma verisiyle karşılaştırılıp, konuşmanın kime ait olduğunu bulma işlemidir. Özniteliklerden kimliklerin oluşturulması konuşmacı tanıma sistemlerinin en kritik aşamasıdır. Bu aşamada oluşturulan kimliğin iki temel özeliği sağlaması çok önemlidir. Bunlar dayanıklılık ve algılanabilir benzerliktir [1]. Dayanıklılık, oluşturulan kimliğin herhangi bir zamanda alınan ve değişik bozunumlara uğramış konuşma verisinden elde edilen kimliğe büyük oranda benzemesidir. Algılanabilir benzerlik ise farklı zamanlarda aynı kişiden alınan kayıtlardan elde edilen kimlik grubunun, diğer konuşmacılardan elde edilen kimlik gruplarından belirgin bir şekilde ayrılabilir olmasıdır [1]. Kısacası, bir konuşmacı tanıma sisteminin konuşmacı içi değişimi düşük, konuşmacılar arası değişimi yüksek bir modele sahip olması istenir [1]. Ayrıca kurulan modellerin hızlı bir şekilde çalıştırılabilmesi amacıyla boyutlarının orijinal ses kaydına göre çok düşük olması, kurulan konuşmacı tanıma sisteminin hızının belirlenmesinde kritik rol alır. Bu noktada bir ikilem mevcuttur. Modelin boyutlarını büyük tutup dayanıklılık ve algılanabilir benzerliği iyileştirme ile küçük model boyutuna sahip hızlı bir konuşmacı tanıma sistemi tasarlama bir çelişki

oluşturmaktadır. Burada kritik olan nokta, model parametrelerinin etkin bir şekilde belirlenmesi ile küçük model boyutları kullanarak yüksek performans elde edilmesidir [2]. Konuşmacı tanıma sistemleri kullanılan metnin durumuna göre, metinden bağımsız konuşmacı tanıma ve metne bağımlı konuşmacı tanıma olarak ikiye ayrılır [2]. Metne bağımlı konuşmacı tanıma test edilecek konuşmacının hangi metni okuyacağı bilinir. Ancak metinden bağımsız konuşmacı tanıma sistemlerinde böyle bir zorunluluk yoktur. Metne bağımlı konuşmacı tanıma sistemleri daha çok bir yere erişimin kontrollü olduğu uygulamalarda tercih edilen bir yöntemken, adli uygulamalarda daha çok metinden bağımsız konuşmacı tanıma sistemleri kullanılmaktadır.

Kapalı küme ve açık küme, konuşmacı tanıma sistemlerinde görülen iki farklı uygulamadır [2]. Kapalı kümede gelen test verisinin, daha önceden sisteme tanıtılan kişilerden birine ait olduğu bilinir ve sistem test kişisini benzerliği maksimum olan kişiye eşler. Ancak açık küme konuşmacı tanıma sisteminde gelen test verisinin sisteme tanıtılmış olan kişilerden biri olma zorunluluğu yoktur. Yani, sistem için fazladan bir durum daha oluşmaktadır. Bu durum, gelen test verisinin sisteme tanımlanmış biri olmadığı durumudur. Bu durumda sistemin bu ayrımı yapabilmesi için bir sınır değer belirleme gerekliliği doğar [2].

Konuşmacı tanıma probleminin henüz %100 çözülmemiş olmasının altında yatan en büyük nedenlerden biri, insan sesinin yaş, sigara ve alkol kullanımı, hastalık vb. nedenlerden dolayı değişmesi gösterilebilir [2]. Bu açıdan, güvenlik sistemlerinin tasarımında çok fazla kullanılamamaktadır. Bu amaçla kurulan sistemin daha gürbüz olması adına öznitelik çıkarma ve sınıflandırma algoritmalarının ne kadar güçlü olduğunun yanında, seçilecek anahtar kelimenin de çok önemli olduğu açıktır. Her dil için ayrı yapılması gereken çalışmalarla anahtar kelimelerin seçimine bir disiplin getirilebilir.

Aynı zamanda bir iletişim kanalından geçmiş ses üzerinde uygulanan bir konuşmacı tanıma sistemi mikrofona kaydı ile yapılmış konuşmacı tanıma sistemi ile aynı başarıyı göstermeyebilir. Bu yüzden konuşma sinyalinin kodlanması, iletişim kanalından geçilip kod çözülmesi sonrasında yapılacak konuşmacı tanıma sistemlerinin başarıyı doğal olarak daha düşük olur. Bu amaçla klasik yöntemlerin yanında alternatif yöntemlerin de geliştirilmesinde fayda olduğu aşikârdır [3].

Bu tezde oluşturulan konuşmacı tanıma sisteminin Türkçe kelimeler için performans analizi yapılabilmesi adına, 16 farklı kişiden 48 farklı kelime 6 tekrarla alınan konuşma kaydı 16 KHz örnekleme frekansında depolanmıştır. Kayıtlar ses yalıtımının olduğu profesyonel bir stüdyoda her bir kişi için farklı sayıda oturumda alınmıştır. Böylece, kişinin sesinin zaman içinde değişiminin etkisinin de eğitim ve test sürecinde sisteme dahil olması sağlanmıştır.

Tezin temelini oluşturmak için, bir konuşmacı tanıma sistemi oluşturulmuştur. Bu sistemde kullanılan özniteliklerin zaman ekseninde eşlenmesi amacıyla Dinamik Zaman Eğirme (Dynamic Time Warping - DTW) kullanılmış ve sınıflandırma için Öklid mesafesi kullanılması ile kapalı küme konuşmacı tanıma sistemi tasarlanmıştır. Bu sistemin eğitimi için, veri tabanındaki kayıtların üçte ikisi kullanılmıştır. Geriye kalan üçte birlik kısım test verisi olarak ayrılmış ve sistem başarımı farklı öznitelikler için ayrı ayrı gözlemlenmiştir.

Geliştirilen konuşmacı tanıma sisteminde, ilk olarak MFCC öznitelikleri kullanılmış ve kelime tabanlı konuşmacı tanıma sisteminin performansı ölçülmüştür. Veri tabanında bulunan tüm kelimelerin MFCC özniteliklerinin kullanılmasıyla elde edilen başarımlar incelenmiştir.

Tezin sonraki bölümünde aynı veri tabanı kullanılarak MELP kodlanmış konuşma sinyalinin kod açıldıktan sonra MFCC öznitelikleri ile sınıflandırması yapılmış ve oluşturulan konuşmacı tanıma sisteminin kodlamaya karşı dayanıklılığı konuşmacı tanıma başarımı açısından incelenmiştir. Mikrofon kayıtları kullanılarak yapılan sınıflandırma başarısının ortalama %95,51'den %59,77'a düştüğünün tespiti ile yeni bir yöntem önerilmiştir. Bu yöntemde MELP kodlama yapıldığında elde edilen özniteliklerinden biri olan çizgisel spektrum frekansları (Line Spectral Frequencies – LSF) indisleri kullanılmıştır. LSF indisleri kullanılarak MELP kod kitabından LSF değerleri elde edilmiş ve bu LSF değerleri sınıflandırmaya tabii tutulmuştur. Bu yöntemle konuşma sinyalinin kodlamanın açılıp MFCC uygulandığı durumdaki başarımına göre %10,74'lük bir artış gözlemlenmiştir. Aynı zamanda, öznitelik edinimi için hesaplama yükü gerektirmeyen, sadece dört adet kod kitabı indisindeki 10 birimlik vektörlerin toplanması işlemi ile sınıflandırma işlemine tabi tutulacak hızlı bir sistem oluşturulmuştur.

Tezin son bölümünde MFCC öznitelikleri kullanılarak oluşturulan konuşmacı tanıma sisteminin başarımı anahtar kelime seçimine göre tekrar değerlendirilmiştir. Yapılan analizler sonucunda, sesletimi kolay, yabancı kökenli olmayan sözcüklerde konuşmacı tanıma başarımının, diğer sözcüklere göre daha yüksek olduğu gözlenmiştir.

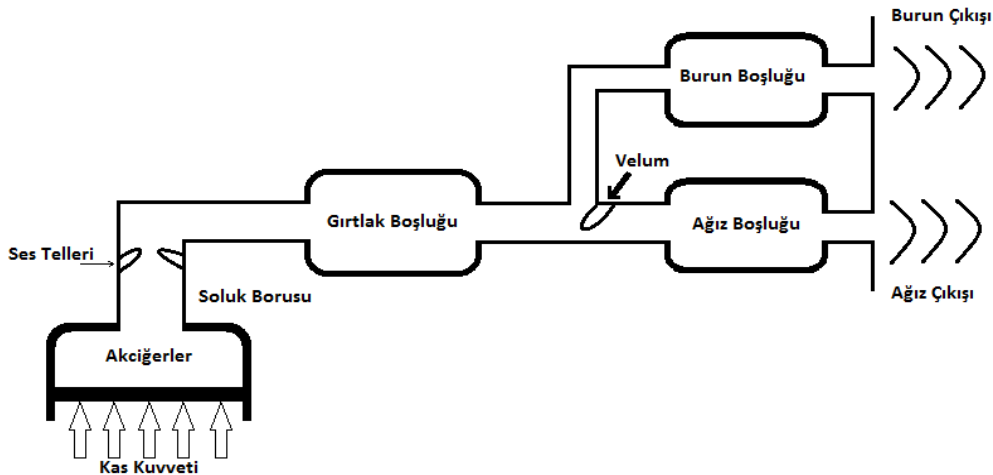
2. KURAMSAL TEMELLER

Bu bölümde tez çalışmasında kullanılmış olan konuşmacı tanıma yöntemlerinin kuramsal temelleri açıklanmıştır.

2.1. Konuşma Sinyalinin Oluşumu

Ses dalgası akustik bir basınç dalgasıdır. Konuşma sinyali ise ciğerlerden üflenlen havanın insanın ses üretim mekanizmasından geçmesiyle oluşur. İnsan ses üretim mekanizmasında temel olarak ses telleri, damak, dişler ve dudak vardır. Ses yolu gırtlak çıkışından başlayıp, dudaklarda sona erer (Selen 1979) ve bu yol boyunca konuşma sinyali şekillenir. Yani, konuşma, tamamen ses yolu üzerinde bulunan anatomik yapıların konum ve duruşlarına göre oluşur. Bu çeşitlilik farklı seslerin ortaya çıkmasını sağlayan temel öğedir. Akciğerlerden çıkan hava gırtlaktan geçerken ses tellerinde oluşturduğu titreşimler çok zayıf sesler oluşturur. Sonrasında ses yolu üzerindeki diğer etmenlerle ses esas formuna ulaşır ve konuşma sinyali üretimi tamamlanmış olur. Ses üretim yolu uzunluğu kadınlarda yaklaşık 15 cm, erkeklerde ise yaklaşık 17 cm civarındadır [4].

Konuşmaların birbirinden ayrılmasının temelinde ses üretim mekanizmasında gerçekleşen olaylar rol almaktadır. İnsan ses üretim mekanizmasının matematiksel olarak iyi modellenmesi konuşma işleme için önemlidir. Şekil 2.1’de insan ses üretim mekanizması genel hatlarıyla gösterilmiştir.

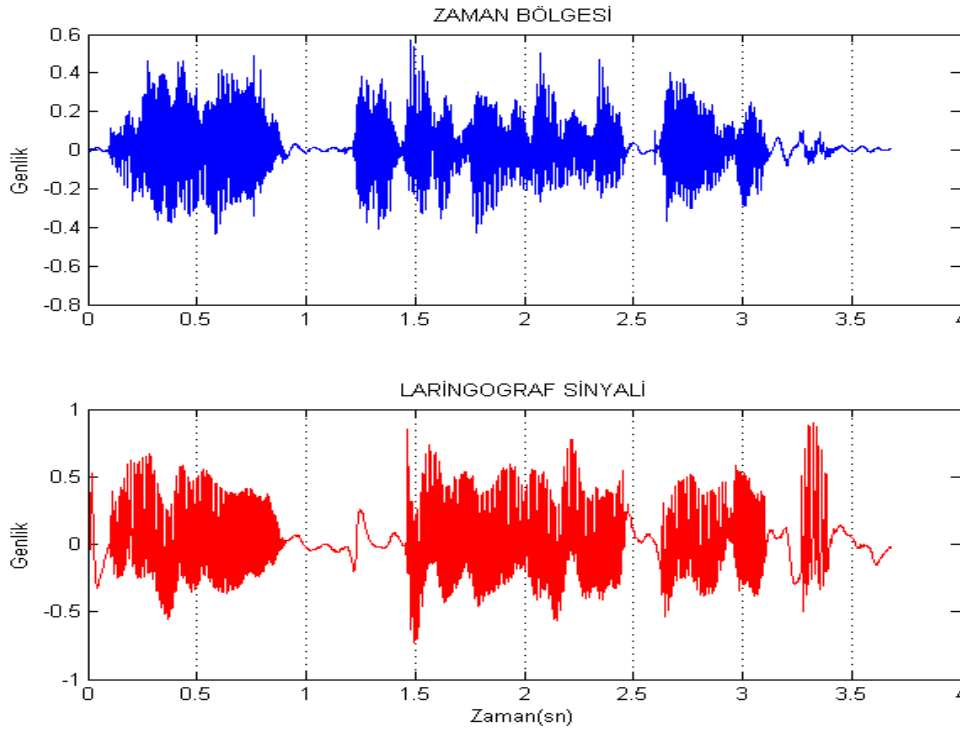


Şekil 2.1. İnsanın konuşma üretim sisteminin çalışması

2.2. Konuşma Sinyalinin Özellikleri

2.2.1. Ötümlü ve ötümsüz sesler

İnsan konuşma sisteminde ses üretimi temelde iki şekilde olur. Bu durumlardan ilki, konuşma sinyali üretilirken ses tellerinin titreşmesi durumudur. Bu türde üretilen ses sinyali hemen hemen periyodik bir yapı gösterir. Bu tarz bir yapı gösteren ses bölütlerine ötümlü ses denir. Diğer durumda ise konuşma sinyalinin üretiminde ses telleri titreşmez ve açık durur. Bu durumda akciğerlerden gelen hava gürültü benzeri bir yapıya sahip olur ve ötümsüz ses oluşur. Konuşma işlemede, sinyalin ötümlü - ötümsüz ayrımının doğru yapılması çok önemlidir. Çünkü insan ses üretim mekanizmasını bir sistem olarak ele alırsak, ses tellerinin titreşip titreşmeme durumu sistem parametrelerinde çok ciddi değişiklik oluşturur. Dolayısıyla konuşma sinyallerinin modellenmesinden önce bu ayrımın doğru yapılması, bundan sonra yapılacak tüm işlemlerin doğruluğu açısından kritik öneme sahiptir. Ötümlü-ötümsüz ayrımı için geliştirilen birçok yöntem vardır [5-8]. Şekil 2.2’de ‘Rana sarı bir zambak giyecek mi?’ cümlesinin iki kanallı kaydı verilmiştir.



Şekil 2.2. ‘Rana sarı bir zambak giyecek mi?’ zaman bölgesi genlik ve Laringograf sinyali gösterimi

İlk kanalda konuşma sinyalinin mikrofonla yapılan zaman bölgesi kaydı, ikinci kanalda ise eş zamanlı Laringograf kaydı verilmiştir. Laringograf, ses tellerinin titreşimlerini boğaza yapıştırılan elektrotlar ile ölçen bir cihazdır [9]. Dolayısıyla Laringograf sinyallerinin periyodik olarak gözlemlendiği bölgeler ötümlü seslerin üretildiğini göstermektedir. Ses tellerinin titreşmediği durumda Laringograf sinyalleri, Şekil 2.2’de 1, 2,5, 3,2 ve 3,5. saniyelerde olduğu gibi periyodik olmayan bir yapı gösterir.

2.2.2. Perde (Pitch) periyodu

Perde periyodu, insan ses üretim sisteminde ses tellerinin titreşim periyodudur. Başka bir deyişle ses tellerinin ardışık iki kapanması arasında geçen süredir. Bir konuşma bölütünün perde periyoduna sahip olabilmesi için ötümlü bir yapıya sahip olması gerekir. Aksi halde ötümsüz bir ses bölütünün perde periyodundan bahsedilemez. Çünkü ötümsüz ses doğası gereği periyodik yapıya sahip değildir. Perde periyodu çocuk, yetişkin erkek ve yetişkin kadın seslerini ayırt etmede tek başına kullanılabilecek bir öznelilik iken, konuşmacı tanıma uygulamalarında tek başına yeterlilik sağlayamamaktadır. Konuşma tanıma uygulamalarında ise çok nadir olarak kullanılır (Rosenberg ve Sambur 1975). Perde periyodunun bulunması için birçok yöntem önerilmiştir [10-13].

2.2.3. Formant frekansları

Ses spektrumunun spektral tepe yaptığı yerler formant frekansı olarak adlandırılır [14]. Ses yolunun rezonans frekansları olarak da adlandırılan formant frekansları, aslında ses yolunun bazı frekanslarının bastırılıp bazılarının vurgulanması ile ortaya çıkar. Bu bakış açısıyla ses yolu bir süzgeç olarak düşünülürse, süzgeçten geçirilmiş sinyalin tepe yaptığı frekans değerleri formant frekanslarını verir. Formantlar, çok önemli akustik parametrelerden biridir. Formant frekanslarının belirlenmesinde literatürde birçok yöntem tanımlanmıştır (Schaefer ve Rabiner 1970).

2.3. Konuşma Sinyalinin Modellenmesi

2.3.1. Doğrusal öngörü katsayıları (Linear prediction coefficients-LPC)

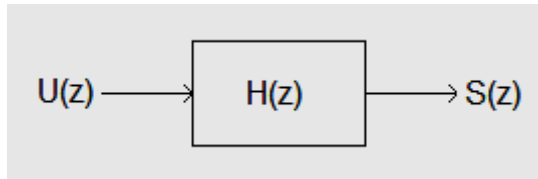
LPC yöntemi [15] konuşma sinyalinin analiz ve sentez işlemlerinde sıklıkla kullanılan etkin bir yöntemdir. Bu yöntemin temel prensibi sinyalin belli bir andaki değerinin kendisinden önceki p tane örneğinin doğrusal kombinasyonu olarak üretilmesine dayanmaktadır (Eş.2.1.).

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (2.1)$$

Eş.2.1’de seçilen p değeri analiz derecesi, a_k ise k . kestirim katsayısı olarak isimlendirilir. Ses yolundan gelen tahrik sinyali (ciğerlerden üflenen havanın modeli) $u(n)$ ise, G kazanç katsayısı ile ölçeklenerek hesaba katılır. Bu kestirimin yapılabilmesi için insan ses üretim yolu sayısal bir süzgeç olarak modellenirse bu sistemin transfer fonksiyonu Eş.2.1’deki gibi olacaktır ve bu sistemin blok şeması Şekil 2.3’de verilmiştir.

$$H(z) = \frac{S(z)}{U(z)} \quad (2.2)$$

Eş.2.2’de, $S(z)$ konuşma sinyalinin z dönüşümü, $U(z)$ ise tahrik sinyalinin z dönüşümü olarak elde edilir.



Şekil 2.3. Konuşma sinyali oluşumunun sayısal süzgeç benzetimi

Bu eşitlikte, Eş. 2.1 yerine yazılırsa insan ses üretim yolu süzgeci Eş.2.3’deki gibi olur. Eş.2.3’deki eşitlikten zaman bölgesine geçilirse Eş.2.1. tekrar elde edilmiş olur.

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{A(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (2.3)$$

$H(z)$ süzgeci matematiksel ifadesinden görüleceđi üzere sadece kutuplardan oluřan bir süzgeçtir ve tahrik sinyalindeki frekans deđerlerinin bir kısmını bastırıp bir kısmını kuvvetlendiren ses iletim yolunu sayısal olarak modellemektedir.

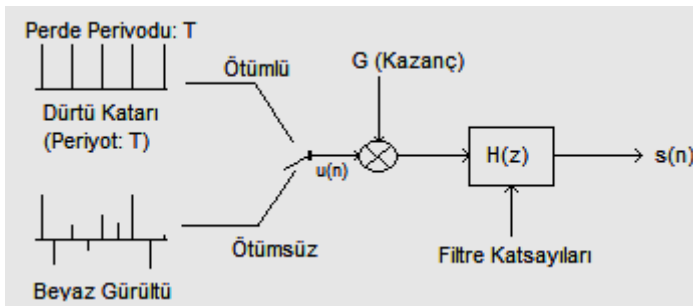
LPC analizi, verilen konuřma sinyalinden $(s(n))$, süzgeç katsayılarının belirlenmesi işleimidir. Katsayıların belirlenmesinde yapılması gereken ise gözlenen sinyal ile asıl sinyal arasındaki farkın minimum olmasının sağlanmasından ibarettir.

$$e^n = s^n - \hat{s}_n = s^n - \sum_{i=1}^P a_i s_{n-i} \quad (2.4)$$

Eř.2.4’de asıl konuřma sinyali s_n ile süzgeç çıkışında elde edilen konuřma sinyali arasındaki fark hata sinyalini vermektedir. Hatanın minimum olması için farklarının karelerinin toplamının minimum olduđu durumu sağlayan katsayılar belirlenmelidir. Bu katsayılar istenen LPC analiz süzgeç katsayıları olacaktır (Eř.2.5).

$$E = \sum_{n=0}^{N-1} e_n^2 = \sum_{n=0}^{N-1} (s_n - \sum_{k=1}^P a_k s_{n-k})^2 \quad (2.5)$$

Bölüm 2.2.1’de ötümlü ve ötümsüz ses kavramından bahsedilmiřti. Ses sinyalinin sentezinde ses tellerinin titreřip titreřmemesi sistem parametrelerini deđiřtiren kritik bir durumdur. LPC sentez işleminde, üretilecek sesin ötümlü olması durumunda $H(z)$ süzgecinin girişine verilecek olan tahrik sinyali, periyodu perde periyodu kadar olan periyodik bir sinyal iken, ötümsüz seslerin sentezinde periyodik olmayan gürültü benzeri (genellikle de beyaz gürültü) tahrik sinyalleri verilir (řekil 2.4).



řekil 2.4. Konuşma sinyali LPC sentezi modeli

2.3.2. Çizgisel spektrum frekansları (Line Spectrum Frequencies – LSF)

LPC parametreleri etkin bir şekilde hesaplanabilir ancak zayıf ara değerlendirme ve nicemleme özelliklerine sahiptir [16]. Bundan dolayı LPC parametrelerinin alternatif şekilde tanımlanacağı yöntemler geliştirilmiştir. Bazı konuşma sentezleme ve konuşma kodlama işlemlerinde bu alternatif yöntemler kullanılmıştır [17-18]. Bu alternatif gösterim yöntemlerinden biri de LPC katsayıları ile eşdeğer gösterim sağlayan çizgisel spektrum frekanslarıdır (LSF). LSF'ler ve formant frekansları arasında çok sıkı bir bağ vardır ve konuşmacının kendine has özelliklerini içerdiklerinden çok sık kullanılırlar [19]. Ayrıca elverişli ara değerlendirme özellikleri ve kararlı olmaları da bir avantajdır [20]. MELP konuşma kodlama algoritmasının ses yolu süzgeç cevabının kodlanmasında LSF'ler kullanılır [21].

LSF'lerin formülleştirilmesi Itakura tarafından gerçekleştirilmiştir [22]. LSF'ler LPC analiz süzgecinden (Eş 2.6) yola çıkarak hesaplanır.

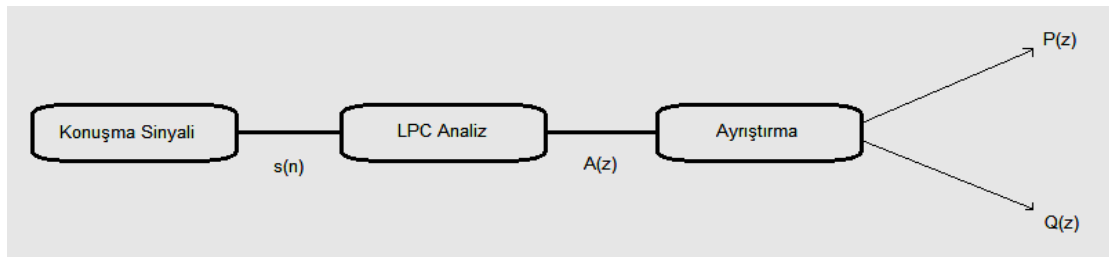
$$A(z) = 1 - \sum_{k=1}^P a_k z^{-k} \quad (2.6)$$

Bu süzgecin derecesini herhangi bir bilgi eklemeyen [23] bir arttırdığımızı $(P+1)$ varsayarsak, kestirim hata süzgecinin polinomları Eş. 2.7 ve Eş 2.8'in toplamı şeklinde (Eş. 2.9) ifade edilebilir (Şekil 2.5).

$$P(z) = A(z) + z^{-(P+1)} A(z^{-1}) \quad (2.7)$$

$$Q(z) = A(z) - z^{-(P+1)} A(z^{-1}) \quad (2.8)$$

$$A(z) = \frac{1}{2} [P(z) + Q(z)] \quad (2.9)$$



Şekil 2.5. Çizgisel Frekans çiftinin çıkarımı

Burada $P(z)$ ve $Q(z)$ ses tellerinin tamamen açık veya tamamen kapalı olduğu durumlarda ses yolunun kayıpsız akustik tüp gösteriminin transfer fonksiyonlarıdır [23]. LSF parametreleri, $P(z)$ ve $Q(z)$ polinomlarının sıfırları olarak ifade edilir ve bu sıfırlar birim çember üzerinde sıralı bir yapıda yer alır. P . derecen bir LPC analiz birim çember üzerinde karmaşık eşlenik olacak konumlarda $2P$ tane frekans içerir. $0 - \pi$ aralığındaki frekans değerlerini göz önünde tutmak yeterli olur. Bu durumda $P(z)$ ve $Q(z)$ 'nin sıfırları belli aralıklarla birim çember üzerinde yer alırlar. $P(z)$ $\omega_1, \omega_3, \omega_5 \dots, \omega_{P-1}$ frekanslarında köklere sahipken $Q(z)$ polinomu $\omega_2, \omega_4, \omega_6 \dots, \omega_P$ frekanslarında köklere sahip olacaktır. Böylelikle birim çember üzerinde $0 - \pi$ aralığında sıralama $0 < \omega_1 < \omega_2 < \omega_3 \dots \dots < \omega_{P-1} < \omega_P < \pi$ gibi olacaktır.

2.3.3. Kepstrum katsayıları

Konuşma sinyalinin oluşumunun, Bölüm 2.3.1'de bahsedildiği gibi ses yolunun sayısal süzgece benzetimiyle doğrusal bir sistem olarak ele alınabileceği belirtilmişti. Sistemin girişine ses yolundan gelen tahrik sinyali, ses yolundan geçerek konuşmayı oluşturmaktadır. Sayısal bir süzgeç olarak tasarlanan ses yolu, $h(n)$ birim dürtü cevabına sahip bir doğrusal sistem olarak düşünülmektedir. Bu sisteme giriş yapan $u(n)$ sinyali ile $s(n)$ konuşma sinyali oluşmaktadır. Zaman bölgesinde konvolüsyon işlemi ile belirlenen konuşma sinyali $s(n)$, frekans bölgesine geçiş yapıldığında çarpıma işlemine dönüşmektedir (Eş.2.6, Eş.2.7).

$$s(n) = e(n) * h(n) \quad (2.10)$$

$$S(w) = E(w)H(w) \quad (2.11)$$

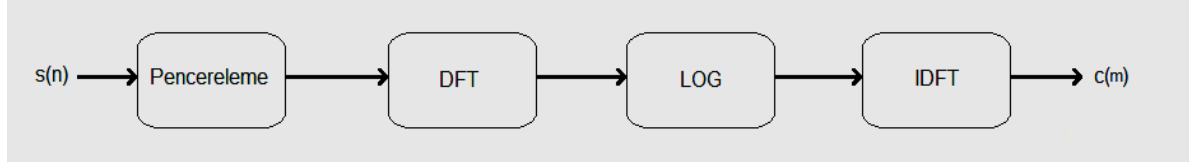
Hem frekans bölgesinde hem de zaman bölgesinde tahrik sinyali ile sistemin dürtü yanıtının ayrıştırılması doğrusal olarak mümkün olmaz. Bu ayrımın doğru ve etkin yapılması birçok konuşma işleme uygulamaları için çok önemlidir. Bu amaçla, denklemin her iki tarafının logaritması alınıp eşitlik toplam formuna getirilebilir (Eş.2.12, Eş.2.13, Eş.2.14).

$$\log(S(w)) = \log(E(w)) + \log(H(w)) \quad (2.12)$$

$$e(n) * h(n) = \log(F(e(n))) + \log(F(h(n))) \quad (2.13)$$

$$c(m) = F^{-1}[\log(F(s(m)))] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|S(w)| e^{jwm} dw \quad (2.14)$$

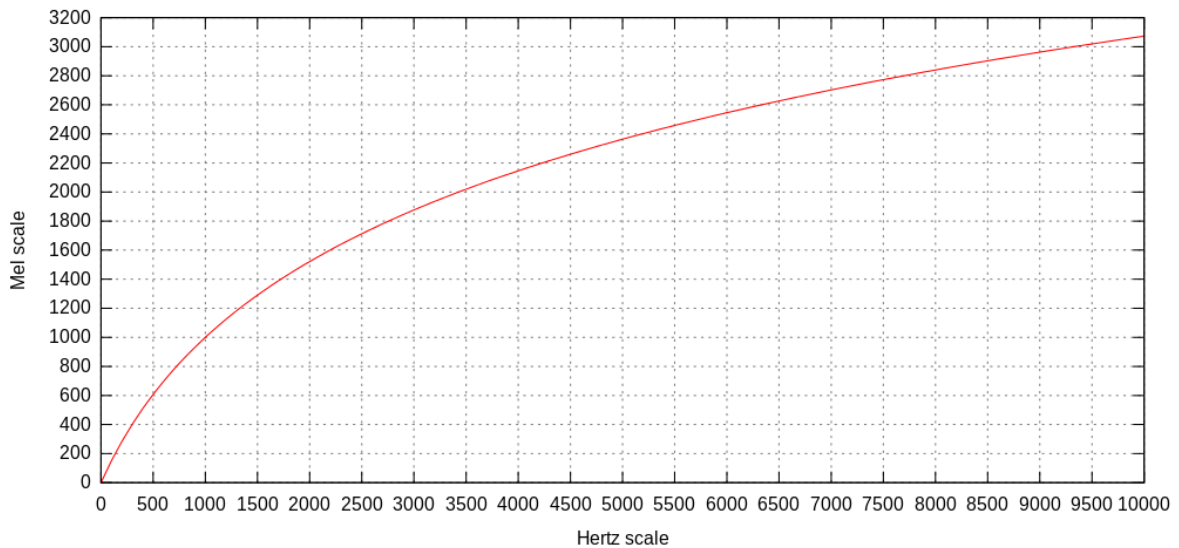
Kepstrum dönüşümü homomorfik dönüşümler sınıfına girmektedir. Kepstral analiz sonucunda elde edilen yeni bölge quefrensy bölgesi olarak isimlendirilir. Tahrik sinyali ve ses yolu süzgeci zamanla değiştiğinden, sinyalin durağan olduğu ve sistemin zamanda değişmez olduğu kısa aralıklar analiz için uygundur (Şekil 2.6).



Şekil 2.6. Konuşma sinyalinin kısa zamanlı kepsral analizi

2.3.4. Mel-frekans kepsrum katsayıları (Mel Frequency Cepstral Coefficients - MFCC)

Konuşma sinyallerinin öznelik çıkarımı için etkin bir şekilde kullanılan MFCC, diğer yöntemlerden farklı olarak insan ses yolunu değil, insan kulağını modellemektedir (Reynolds 1995). İnsan kulağının modellenmesi Stevens ve Volkman (1940) tarafından deneysel olarak yapılmış ve Mel skalası oluşturulmuştur. Bu skalanın oluşturulmasında referans olarak alınan 1000 Hz değerinin karşılığı 1000 Mels olarak seçilmiş ve insan kulağının algıladığı ölçeğe göre Herz-Mels değerleri deneysel olarak belirlenmiştir. Bu deneyin sonucunda Herz-Mels değerlerinin alçak frekans değerlerinde (1 kHz'e kadar) doğrusal olarak eşleştiği, yüksek frekans değerlerinin logaritmik bir değişim olduğu gözlemlenmiştir. Bu eşleşme sonucunda frekans değerlerinin Mel ölçeğinde karşılığı Şekil 2.7'de verilmiştir.

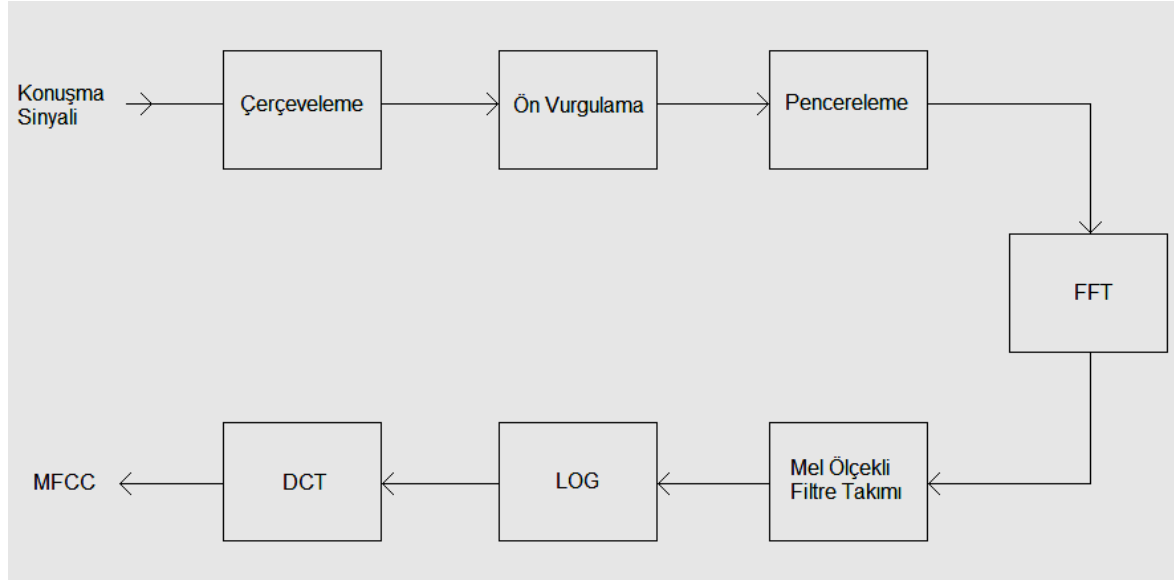


Şekil 2.7. Mel – Frekans dönüşüm eğrisi [24]

Bir frekans değerinin Mel ölçeğindeki karşılığını bulmak için Eş.2.15 kullanılmaktadır.

$$mel(f) = 2595 \log(1 + f/700) \quad (2.15)$$

MFCC öznitelikleri Şekil 2.8’de gösterilen yol izlenerek hesaplanır.



Şekil 2.8. MFCC özniteliklerinin çıkarımı

MFCC çıkarım işleminde, ilk önce konuşma sinyalinin durağan olduğu varsayılan uzunluktaki bölümleri, ardışık pencerelerde bir kısmı çakışacak şekilde seçilir. Bu şekilde çerçeveler arası geçiş ve değişim daha muntazam olur. Sonrasında konuşma sinyali transfer fonksiyonu Eş.2.16’da verilen birinci dereceden sonlu dürtü tepkisi olan (Finite Impulse Response - FIR) bir süzgeçten geçirilir. Ön vurgulama olarak adlandırılan bu işlem, konuşma sinyalinin yüksek frekanslarını daha belirginleştirmek için yapılır.

$$H(z) = 1 - 0,95z^{-1} \quad (2.16)$$

Pencereleme işlemi, frekans bölgesinde analizinden önce, konuşma sinyalinde bulunmayan yalancı frekans değerlerinin ortaya çıkmasını önlemek amacıyla yapılmaktadır. Ortaya çıkan ve aslında olmayan frekans değerleri, pencerenin başındaki ve sonundaki örneklerin farkından kaynaklanmaktadır ve spektral sızma (kaçak) olarak adlandırılmaktadır. Spektral sızmayı azaltmak üzere pencerenin ilk ve son değerlerini bastıran pencereleme fonksiyonları kullanılmaktadır. Değişik pencere fonksiyonları mevcut olmakla birlikte,

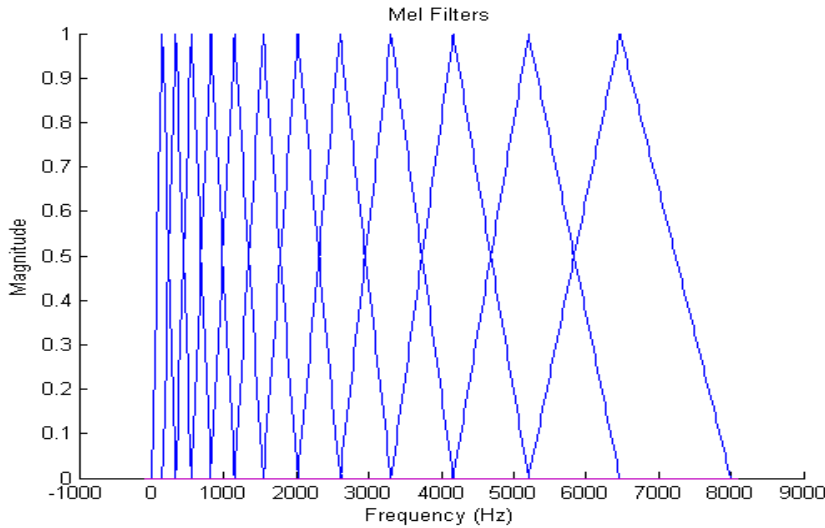
konuşma işlemede matematiksel ifadesi Eş.2.17’de verilen Hamming pencere tercih edilmektedir.

$$w(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 1 \leq n < N \quad (2.17)$$

Deneysel olarak oluşturulan Mel ölçeği süzgeç takımı, 1 kHz’e kadar doğrusal, 1 kHz’den sonra logaritmik olarak yerleştirilen üçgen süzgeçlerden oluşmaktadır ve genlik spektrumu çıkarılan konuşma sinyaline uygulanmaktadır. Şekil 2.9’da k adet süzgeç ile gerçekleştirilmiş bir Mel süzgeç örneği verilmiştir. Her bir süzgeç komşu süzgeçler ile %50 oranında iç içe olacak şekilde konumlandırılmaktadır. Kullanılan süzgeç sayısı kadar MFCC özniteliği elde edilir.

Mel süzgecinden sonra, her bir süzgeç çıkışındaki değerin logaritması alınır. Bu şekilde değişimlerden çok fazla etkilenmeyecek şekilde korunmuş olur. Süzgeçlerden herhangi birinin logaritmik enerji çıkışı $c(p)$ Eş.2.18’teki gibi hesaplanmaktadır.

$$c(p) = \log\left(\frac{1}{A_p} \sum_{k=L_p}^{J_p} F_p[k]X[k]\right) \quad (2.18)$$



Şekil 2.9. MFCC öznitelik çıkarımı için kullanılan süzgeç takımı

MFCC değerlerinin çıkarımında son aşama Mel spektrum katsayılarının ayırık kosinüs dönüşümü (Discrete Cosine Transform - DCT) ile tekrar zaman alanına geçilmesidir. MFCC katsayıları Eş.2.19 yardımı ile hesaplanır.

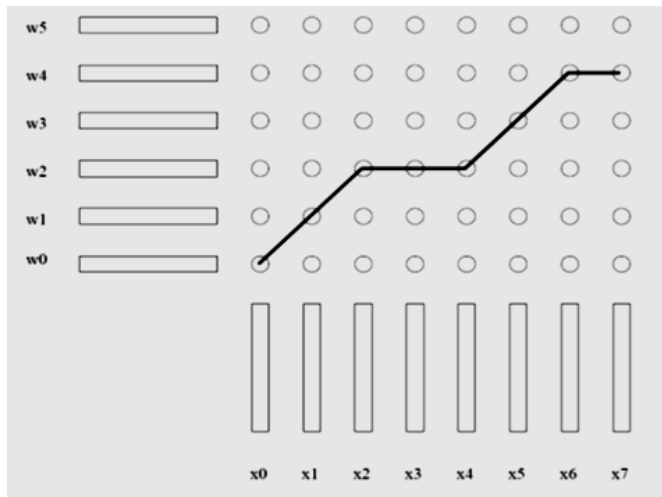
$$MFCC(j) = \frac{1}{f_s} + \sum_{i=1}^{f_s} c_i \cos\left(j\left(1 - \frac{1}{2}\right)\frac{\pi}{f_s}\right), \quad j = 1, 2, \dots, (f_s - 1) \quad (2.19)$$

2.4. Konuşma Sinyalinin Sınıflandırılması

Sınıflandırma işlemi, bir konuşmacı tanıma sisteminin eğitimi aşamasında her bir kişi için çıkarılan özneliklerin, test verileri ile olan ilişkisinin belirlenmesi sürecidir. Konuşmacı tanıma uygulamalarında karşılaşılan birçok sınıflandırma uygulamaları mevcuttur [25-27].

2.4.1. Dinamik zaman eğirme (Dynamic Time Warping - DTW)

Aynı konuşmacının aynı kelimeyi her söyleyişi, aynı ortam ve şartlar altında olsa dahi, daha önceki söyleyişlerinden farklı olmaktadır. Metne bağlı konuşmacı tanıma işleminde aynı kelimelerin farklı kayıtlarının aynı uzunluğa getirilebilmesi için kullanılan şablon modellerden biri DTW'dir. DTW'de bir referans kayıt alınıp, diğer kayıtların boyları referans kayda göre yeniden düzenlenir. Bu şekilde her bir sınıfın kayıtlarından elde edilen öznelik vektörü boyutu eşit olur. DTW yöntemiyle tasarlanmış metne bağlı konuşmacı tanıma uygulamaları mevcuttur [2].



Şekil 2.10. Aynı niteliklere sahip farklı uzunluktaki iki verinin eşleştirilmesi [28]

Şekil 2.10, aynı kelimenin x ve w ile simgelenen iki farklı zamanlarda aynı kişi tarafından ses kayıtlarının eşlenmesini göstermektedir. Kayıtlardan biri 6 pencere boyutuna sahipken diğeri 8 pencere boyutundadır. DTW algoritması ile iki kaydın pencerelerini sırasıyla karşılaştırıp en benzer olanları birbirine eşlenir. Şekil 2.10'da x_0 penceresi w_0 'a, x_1

penceresi w_1 'e, x_2 penceresi w_2 'ye, x_3 penceresi tekrar w_2 'ye ... şeklinde eşleştirilmiştir ve eşleştirme yolu P şeklindeki gibi belirlenmiştir. Bu eşleştirme işleminde istatistiksel olarak çok fazla ihtimal olmasının yanında, P yolunun herhangi bir anda daha önceki pencerelere eşlenmesi ihtimal dışıdır. Çünkü söylenen kelimeler aynı olduğundan ve eşleştirme kelime başından sonuna doğru olacağından, P eğrisi bir sonraki adımda ya önceki adımla aynı pencereye, ya da sonraki pencereye eşlenecektir. Bu yolla ideal P eğrisi belirlenir ve veriler, dolayısıyla öznitelik boyutları eşitlenmiş olur.

2.4.2. Vektör nicemleme (Vector Quantization - VQ)

En yakın komşuluk algoritmasını kullanarak verilerin sınıflara ayrılmasını amaçlayan VQ tekniği, konuşmacı tanıma algoritmalarında sıklıkla kullanılan bir yöntemdir. Bu yöntemde, N konuşmacının her birinden elde edilen M adet öznitelik vektörü 2^k tane vektör ile temsil edilir. Burada, k değerinin sıfır seçilmesi, M adet öznitelik vektörünün ortalamasının alınmasına karşılık gelir ki bu da her bir konuşmacının bir adet kod kelimesi (centroid) ile temsil edilmesi demektir. N konuşmacı için de toplamda N adet kod kelimesine sahip bir kod kitabı oluşturulmuş olur. k 'nın daha büyük değerleri için her bir konuşmacıyı temsil edecek vektör sayısı artar. Bu vektörlerin bulunmasında genellikle optimum vektör nicemleme algoritması olan LGB (Linde, Bum & Gray, 1980) kullanılır.

Bu tez çalışması kapsamında, DTW kullanılarak sözcüklerin içerdiği pencere sayıları eşitlendikten sonra, karşılaştırma için $k = 0$ seçilerek öznitelikler üzerinde vektör nicemleme yapılmıştır. Literatürde çok sayıda farklı konuşmacı tanıma yöntemi önerilmiştir. Parametrik olmayan yaklaşımla [1], destek vektör makinaları kullanılarak [29], yapay sinir ağları kullanılarak [30], dalgacık dönüşümleri kullanılarak [31], Gauss Karışım modeli kullanılarak [32] yapılan konuşmacı tanıma sistemler bunlar arasında gösterilebilir. Bu tez çalışması kapsamında ise, hızlı ve güvenilir çalışması önceki çalışmalardan bilinen DTW ve VQ yöntemleri temel alınmıştır.

3. KONUŞMACI VERİTABANININ OLUŞTURULMASI

Bu çalışmada kullanılan veri tabanı, 16 farklı kişinin, 48 farklı kelimeyi, farklı zamanlarda altışar kez söylemesi ile oluşturulmuştur. Şekil 3.1’de kayıtların alındığı profesyonel stüdyo mevcuttur.



Resim 3.1. Kayıtların alındığı profesyonel stüdyo

Kayıtların alındığı stüdyo dışarıdan gelen gürültülere karşı yalıtılmıştır. Aynı zamanda duvarlarındaki ses emici süngerler sayesinde konuşma sinyalinin duvarlara çarparak oluşturacağı muhtemel yansımalar engellenmiştir. Resim 3.2 duvarlardaki ses emici yapıları göstermektedir.

Kayıtlarda kullanılan mikrofon Logitech marka kablosuz kafa üstü bir mikrofondur. Veri seti kayıt alımı için Wavesurfer programı kullanılmış ve veriler .wav formatında 16 kHz örnekleme frekansında depolanmıştır. Katılımcıların kayıtlarının alımı esnasında stüdyoda tek başına olması sağlanmıştır. Ayrıca şebeke frekansının dizüstü bilgisayardan mikrofon kayıtlarına girişim yapma ihtimaline karşı dizüstü bilgisayar pille çalışır durumda iken kayıt işlemi gerçekleştirilmiştir.



Resim 3.2. Stüdyo duvarlarındaki yalıtım malzemesi

Veri setinin oluşturulmasında her bir katılımcının aynı şartlarda kayıtlarının yapılmasına azami dikkat edilmiştir. Stüdyo içerisindeki sandalye ve masanın konumlarının aynı kalmasına, mikrofonun yüz üzerindeki konumunun her bir katılımcı için standart olmasına dikkat edilmiştir. Kayıtlar esnasında karşılaşılabilecek sıralı okuma psikolojisi ve diğer gerek psikolojik gerekse de fizyolojik bozucu etmenlerin kayıt öncesinde katılımcıya anlatılmasıyla mümkün olduğunca standart veriler elde edilmesine çalışılmıştır. Şekil 3.3’de kayıt esnasında her bir konuşmacıda mikrofonun konumu gösterilmektedir.



Resim 3.3. Kayıt alımı sırasında baş üstü mikrofonun konumu

Veri setine katkı sağlayan konuşmacılar 12 erkek ve 4 kadından oluşmaktadır. Katılımcıların yaşları 25 ve 40 arasındadır. Hiçbir katılımcı sigara kullanmamaktadır. Katılımcılarla ilgili genel bilgiler Çizelge 3.1’de verilmiştir.

Çizelge 3.1. Veri tabanına katkı sağlayan konuşmacıların bilgileri

	CİNSİYET	DOĞUM YILI	MEMLEKETİ	EĞİTİM DURUMU
1	E	1980	BAYBURT	Y.LİSANS
2	E	1985	ÇORUM	LİSANS
3	E	1988	AFYON	Y.LİSANS
4	E	1987	TRABZON	LİSANS
5	E	1987	ARTVİN	LİSANS
6	E	1986	İSTANBUL	Y.LİSANS
7	E	1985	İSTANBUL	Y.LİSANS
8	E	1989	MALATYA	LİSANS
9	E	1988	NİĞDE	LİSANS
10	E	1986	MALATYA	LİSANS
11	E	1987	İSTANBUL	LİSANS
12	E	1989	SİVAS	LİSANS
13	K	1987	ANKARA	LİSE
14	K	1977	HATAY	LİSANS
15	K	1986	ERZURUM	LİSANS
16	K	1980	ANKARA	LİSANS

Veri setinde kullanılan kelimeler, Türkçedeki farklı ses kombinasyonlarını mümkün oldukça fazla içerecek şekilde seçilmeye çalışılmıştır. Aynı zamanda kelimeler arasında Türkçe 'ye özellikle Fransızca ve Arapça olmak üzere diğer dillerden geçmiş örnekler de bulunmaktadır.

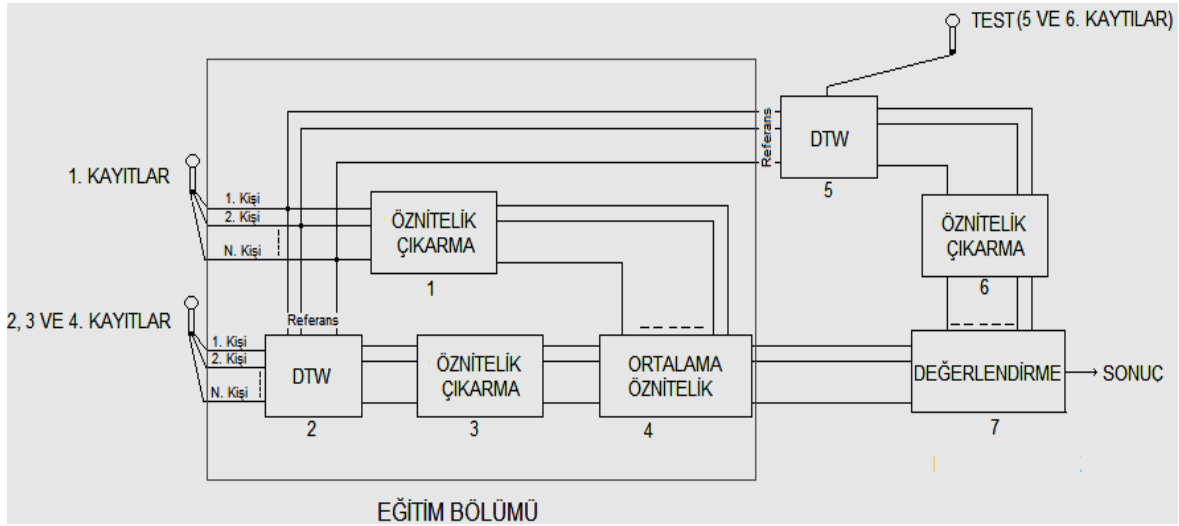
Kayıtlar 16 kHz örnekleme frekansında 16 bit nicemleme seviyesinde yapılmıştır. Veri seti, MELP kodlanmış konuşmacı tanıma sistemi için kullanılırken örnek seyreltme ile 8 kHz örnekleme frekansına dönüştürülmüştür. Tablo 3.2’de veri tabanında kullanılan kelimeler verilmiştir.

Çizelge 3.2. Veri tabanında bulunan 48 farklı kelime ve veri tabanı indisleri

VERİTABANINDAKİ KELİMELER							
1.AKTÖR	7.EKONOMİK	13.KAPAK	19.TUZAK	25.TAM	31.DİLEK	37.KIRAZ	43.GÜNEŞ
2.BARIŞ	8.BEŞİK	14.KARTOPU	20.BOZUK	26.TEKRAR	32.KADIN	38.GARANTİ	44.POLİS
3.TOPRAK	9.GİTAR	15.KIRMIZI	21.BÖREK	27.BÜTÜN	33.DOLAP	39.GENİŞ	45.PÖRSÜK
4.DAMAR	10.GURBET	16.KÖPÜK	22.PAZAR	28.BIÇAK	34.DÖŞEK	40.GOFRET	46.PIRASA
5.DENİZ	11.METİN	17.TIRNAK	23.PERŞEMBE	29.BİDON	35.KEDİ	41.GÖMLEK	47.PİLAV
6.DÜMDÜZ	12.HÜKÜMET	18.TÜTÜN	24.BURUŞUK	30.DUMAN	36.KUZU	42.GIDA	48.PÜRE

4. KONUŞMACI TANIMA SİSTEMİ

Konuşmacı tanıma sistemlerinde amaç, sisteme girdi olarak verilen bir konuşma sinyalinin kime ait olduğunun en kısa sürede ve en güvenilir şekilde belirlenmesidir. Konuşmacı tanıma için geliştirilmiş birçok sistem mevcuttur [1,29-32]. Bu sistemler, kullanılan öznitelik çıkarım yönteminden test verilerini sınıflandırma yöntemlerine kadar birçok noktada farklı yapıda olabilirler. Bu sistemlerde amaç her koşulda %100 başarı sağlanmasıdır. Ancak, kişinin konuşma sinyalinin çok fazla parametreye bağlı olarak değişiklik gösterebileceği gerçeği, tam manasıyla güvenilir bir sistem oluşturulmasının önünde en büyük engel olarak gösterilebilir.



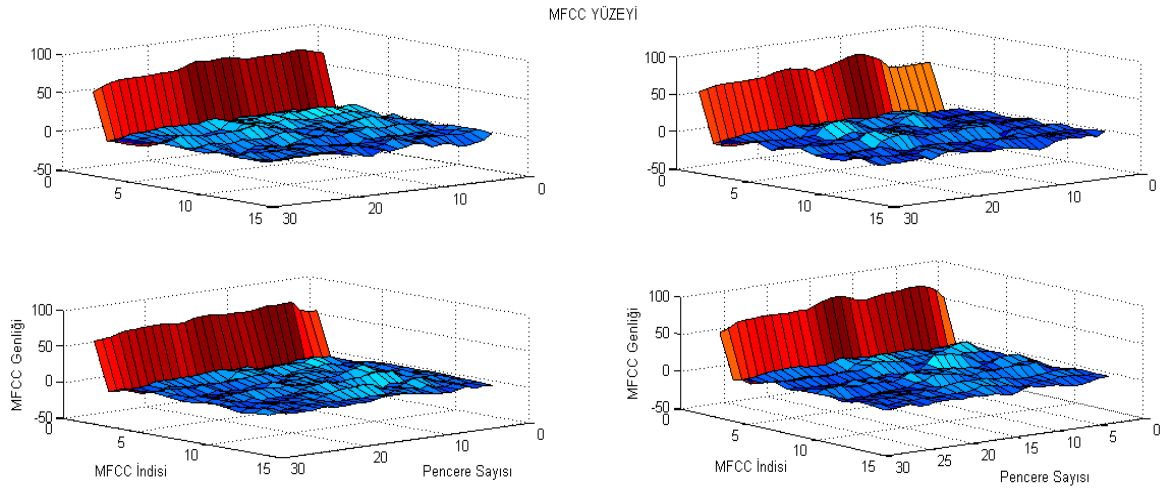
Şekil 4.1. Konuşmacı tanıma sistemi

Bu tez kapsamında oluşturulan konuşmacı tanıma sistemi Şekil 4.1'deki gibidir. Bu sistemin eğitim kısmında, her bir konuşmacı için bir öznitelik vektörü çıkarılır ve kişi sistemde o vektör ile temsil edilir. Öznitelik vektörlerinin boyutlarını her bir konuşmacı için standart hale getirmek amacıyla, ilk kayıtlar referans alınarak diğer kayıtlar DTW yöntemiyle referans kayıt uzunluklarına eşlenir. Şekil 4.1'deki 2. blokta bu işlem gösterilmektedir. Daha sonra öznitelik çıkarma yöntemiyle ilgili öznitelikler eğitim kümesindeki tüm kayıtlardan elde edilip (Blok 1,3), her kişiye ait sadece bir öznitelik vektörü olması için ortalama öznitelik hesabı yapılmıştır (Blok 4). Eğitim işlemi için veri tabanındaki kayıtların üçte ikisi kullanılmıştır. Eğitim işleminin sonucunda her bir konuşmacıya ait öznitelik vektörü test verilerinin karşılaştırılması amacıyla kaydedilmiştir. Sistemin başarı oranının tespitinde %33'lük test kayıtları sırasıyla sisteme verilmiştir. Her

bir test verisi ilk önce blok 5’deki DTW bloğunda sırasıyla tüm konuşmacıların uzunluklarına eşlenir ve her bir eşleme için öznitelik hesabı yapılarak (Blok 6), eşlenen kişiye ait eğitim sırasında elde edilen öznitelik vektörü ile olan Öklid mesafesi hesaplanır (Blok 7). Öklid mesafesinin en küçük olduğu kişi, test verisinin sistem tarafından eşlendiği sonuç olur.

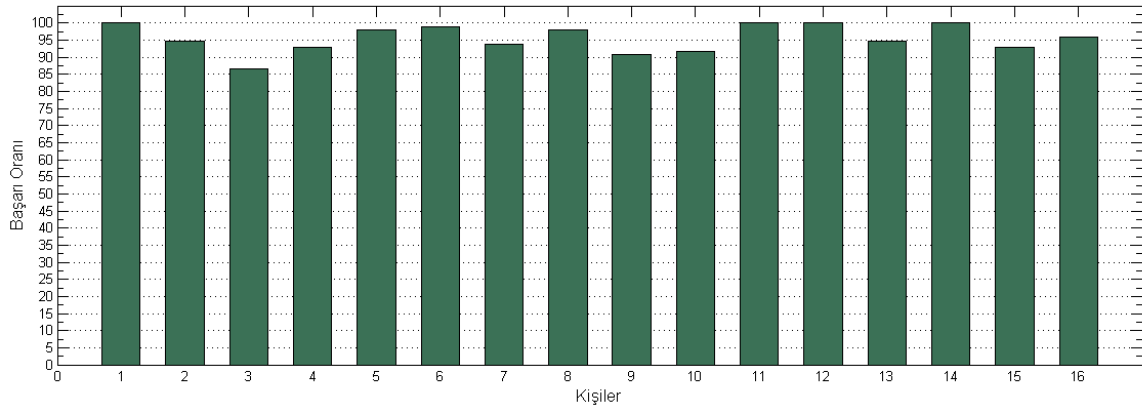
4.1. MFCC Öznitelikleri Kullanılarak Sınıflandırma Başarımı

Mel frekans kepstrum katsayıları (MFCC) insan kulağının duyması taklit edilerek oluşturulmuş bir öznitelik çıkarma yöntemidir. MFCC ile yapılan öznitelik çıkarımı, konuşmacı tanıma uygulamaları için başarılı sonuçlar vermektedir. Şekil 4.1’deki konuşmacı tanıma sisteminin eğitim kısmında her bir kişinin her kelimesi için MFCC öznitelikleri elde edilir ve her bir test verisi için elde edilen MFCC öznitelik değerleri ile karşılaştırılır. Şekil 4.2’de veri tabanının eğitim kısmında elde edilen 4 kişiye ait “dümdüz” sözcüğünün MFCC yüzeyleri verilmiştir.



Şekil 4.2. Veri tabanındaki 4 kişiye ait “dümdüz” sözcüğünün MFCC yüzeyi

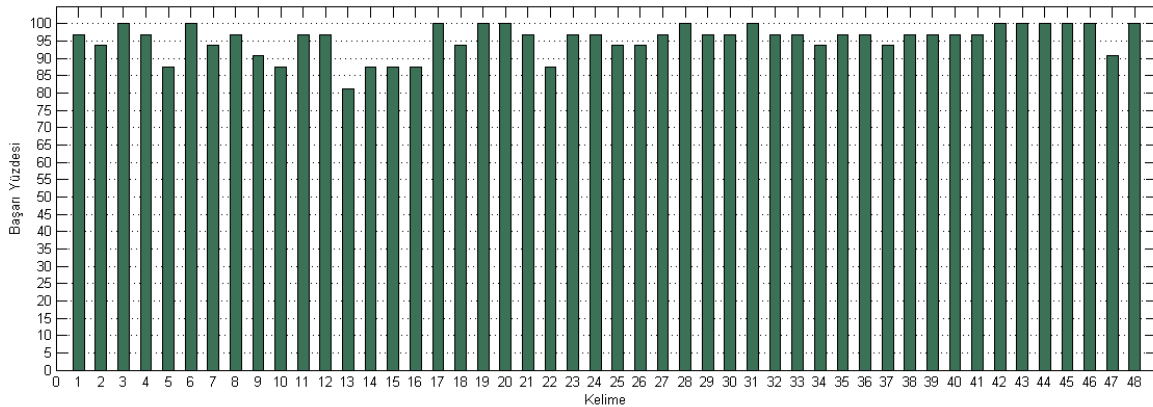
Veri tabanındaki 16 kişinin, MFCC öznitelikleri ile yapılan sınıflandırmadaki yüzde başarımı Şekil 4.3’te verilmiştir.



Şekil 4.3. MFCC öznitelikleri kullanılarak yapılan sınıflandırmada kişilerin yüzde başarıları

4 kişinin tüm kelimeleri sistem tarafından %100 doğruluk ile sınıflandırılmıştır. 3. kişi %86,46 doğru tanınma yüzdesi ile en az orana sahiptir.

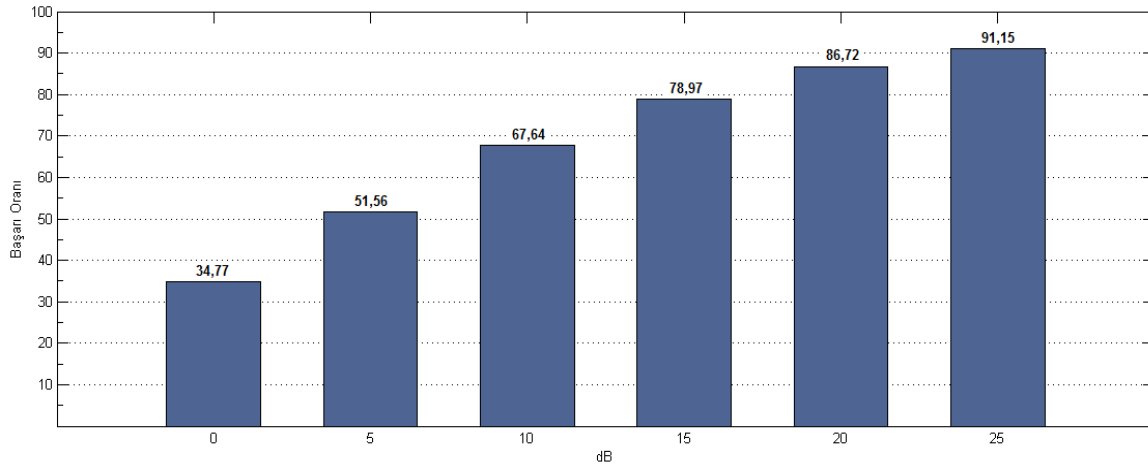
Sistemin kelimelere göre yüzde başarı tablosu Şekil 4.4’te verilmiştir. Bu şekildeki kelime başarıları kişiler üzerinden alınan ortalamalardan oluşmaktadır. Kelimelerin indisleri Çizelge 3.2’de gösterilmiştir. Veri tabanında kullanılan kelimelerden 13’ünün kime ait olduğu sistem tarafından tam olarak tespit edilmiştir. 13. kelime olan “Kapak” kelimesi ise %81,25 ile en az tanınma yüzdesine sahiptir. MFCC öznitelikleri kullanılarak oluşturulan sistemin toplam başarı oranı %95,51 olmuştur.



Şekil 4.4. MFCC öznitelikleri kullanılarak yapılan sınıflandırmada kelimelerin yüzde başarıları

4.1.1. Test verisine farklı desibel (dB) değerlerinde sinyal gürültü oranını (SNR) sağlayacak beyaz Gauss gürültü eklenmesinin sistem performansına etkisi

Tasarlanan sistemler, ortam koşullarından kaynaklı her türlü etkiye maruz kalabilirler. Bu durum, ideal olarak yapılan hesaplama ve gözlemlerden farklı sonuçlara hazırlıklı olmamız gerektiği sonucunu doğurur. Güvenlik için tasarlanan bir konuşmacı tanıma sisteminde de bu etkilerin göz önüne alınarak başarımların bu etkilere göre yorumlanması önem arz eder. Konuşma sinyali göz önüne alındığında, konuşmacının test verisini sisteme verdiği sırada, herhangi bir nedenden kaynaklı arka plan gürültüsü de sistem tarafından algılanabilir. Bu durumda sistem performansını gözlemlemek adına, kurulan konuşmacı tanıma sisteminde test verilerinin her birisine sırasıyla 0, 5, 10, 15, 20 ve 25 desibel (dB) sinyal gürültü oranını (signal to noise ratio - SNR) sağlayacak gürültü eklenmiştir. Şekil 4.5'te sistemin testi için kullanılacak sinyallere, farklı desibellerde sinyal-gürültü oranını sağlayacak beyaz Gauss gürültüsü eklendiğindeki sistem performansı gösterilmektedir.



Şekil 4.5. Sistemin testi için kullanılacak sinyallerin farklı SNR değerlerinde sistem başarımları

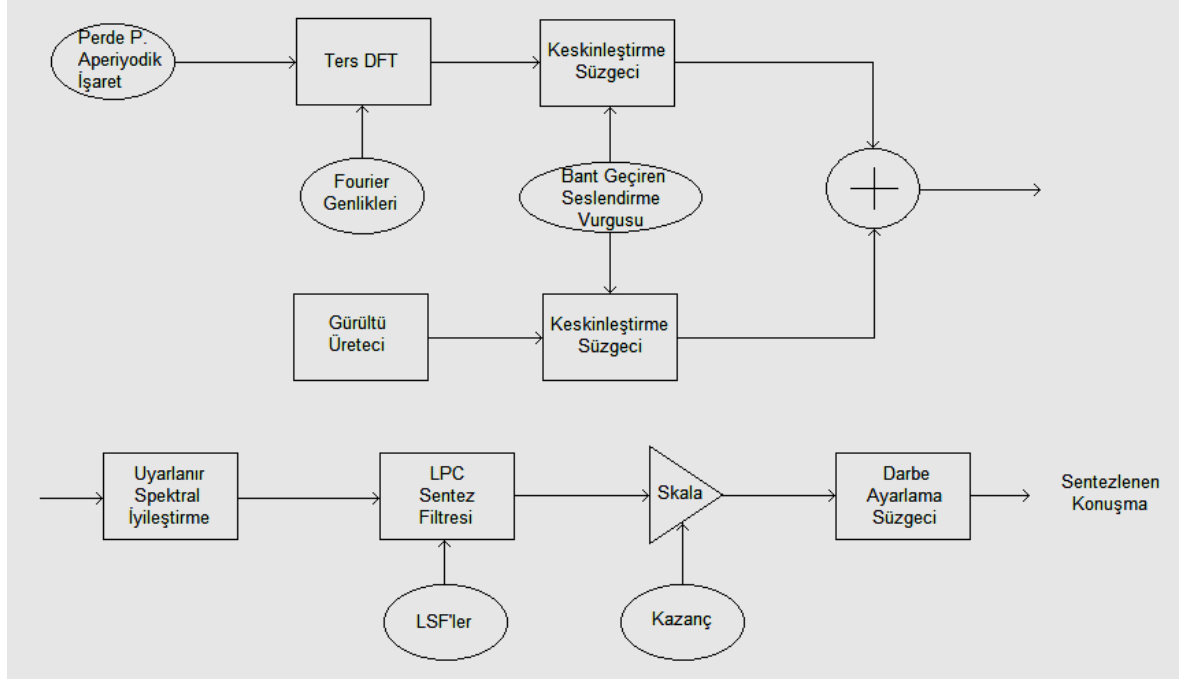
Konuşmacı tanıma sistemlerinin tasarımı, sadece mikrofon kayıtları üzerinden güvenli erişim amacıyla yapılan çalışmalarla sınırlı kalmayıp, kablosuz iletişim hatları üzerinden güvenli ve kontrollü erişimin sağlanabilmesi şeklinde karşımıza çıkabilmektedir. Bu durumda kanalda eklenecek gürültünün yanında konuşma sinyalinin kodlanıp kod açılmasında kaynaklı bozunum ve gürültüler de eklenecektir. Bu etkinin incelenmesi adına yaygın olarak kullanılan MELP konuşma kodlamanın, konuşmacı tanıma performansına etkisi Bölüm 4.2'de tartışılacaktır.

4.2. MELP Kodlanıp Kod Açılmış Veri Tabanıyla MFCC Öznitelikleri Kullanarak Sınıflandırma Başarımı

MELP (Mixed Excitation Linear Prediction), klasik LPC (Linear Prediction Coefficient) parametrik model tabanlı kodlamayı kullanan bir konuşma kodlama yöntemidir. MELP kodlama ile konuşma sinyali saniyede 2400 bitle temsil edilecek bir dönüşüme uğrayarak sayısallaştırılır. MELP'i LPC kodlamadan ayrılan 5 temel özelliği vardır [21]. Bunlar: karışık uyartım, aperiodyik darbe, uyarlanır spektral iyileştirme, darbe dağılımı ve Fourier genlik modellemesidir [21]. Bu özellikleriyle, MELP kodlama, konuşma kalitesi açısından oldukça başarılıdır ve en çok kullanılan konuşma kodlama yöntemlerinden biridir. Bu tez çalışmasında, MELP kodlama, geliştirilen konuşmacı tanıma sisteminin kodlamaya karşı dayanıklılığını denemek için kullanılmıştır.

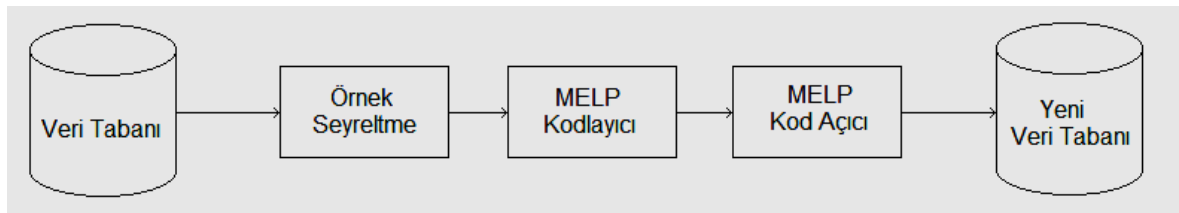
Kodlama sürecindeki adımlar analog giriş sinyalinde çok düşük frekansların silinmesi ile başlar ve sırasıyla işlemler şöyle devam eder: Perde periyodu hesabı, bant geçiren seslendirme analizi, kesirli perde periyodu düzeltmesi, aperiodyik işaretleme, LPC analizi, doğrusal öngörü kalıntı hesabı, son perde periyodu hesabı, perde periyodu ikileme hatasının düzeltilmesi, kazanç hesabı, ortalama perde periyodu güncellemesi, öngörü katsayıları nicemlenmesi, perde periyodu nicemlenmesi, kazanç nicemlenmesi, bant geçiren seslendirme nicemlenmesi, Fourier genlik hesabı ve nicemlenmesi, hata koruması ve bit paketlenmesi [21].

MELP kod açılma süreci, bit paketi açılması ve hata düzeltme adımıyla başlar, gürültü azaltma, parametre ara değerlemesi, karışık uyartım üretimi, uyarlanır spektral iyileştirme, doğrusal öngörü sentezi, kazanç ayarlaması, darbe ayarlaması ve sentez çevrim kontrolü ile kod açılma işlemi tamamlanmış olur. MELP kod açma blok diyagramı Şekil 4.6'da verilmiştir.



Şekil 4.6. MELP kod açıcı blok diyagramı

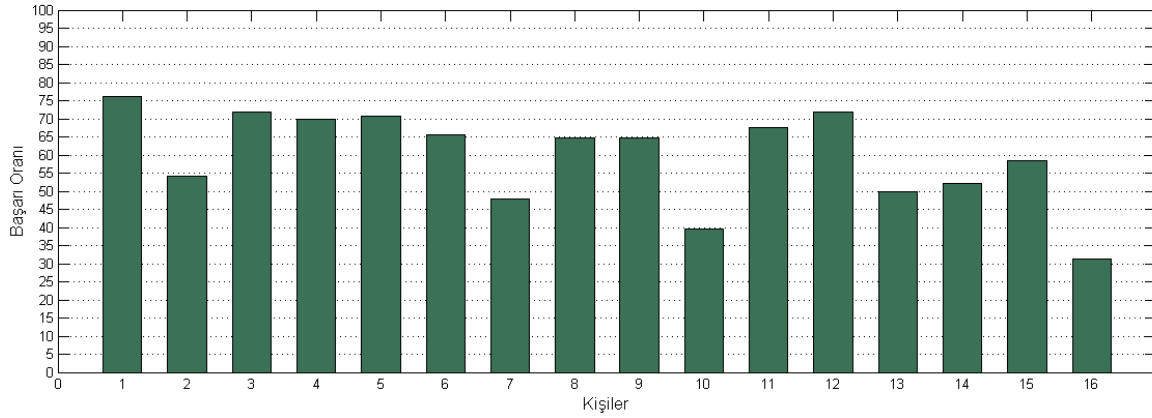
Bu bölümde yapılan işlem Bölüm 4.1'deki işlem aynı olmakla birlikte, kullanılan veri tabanı üzerinde değişiklik yapılmıştır. Burada amaç, kodlanıp kod açılmış sinyalin MFCC özniteliklerine dayalı sistem başarımına etkisinin incelenmesidir. Bölüm 3'te oluşturulan veri tabanının kodlamanın etkisini incelemek adına yeniden düzenlenmesi gerekir. Veri tabanının hangi işlemlerden geçirilip yeni veri tabanının oluşturulduğu Şekil 4.7'de gösterilmiştir.



Şekil 4.7. MELP kodlanıp kod açılmış veri tabanının oluşturulması

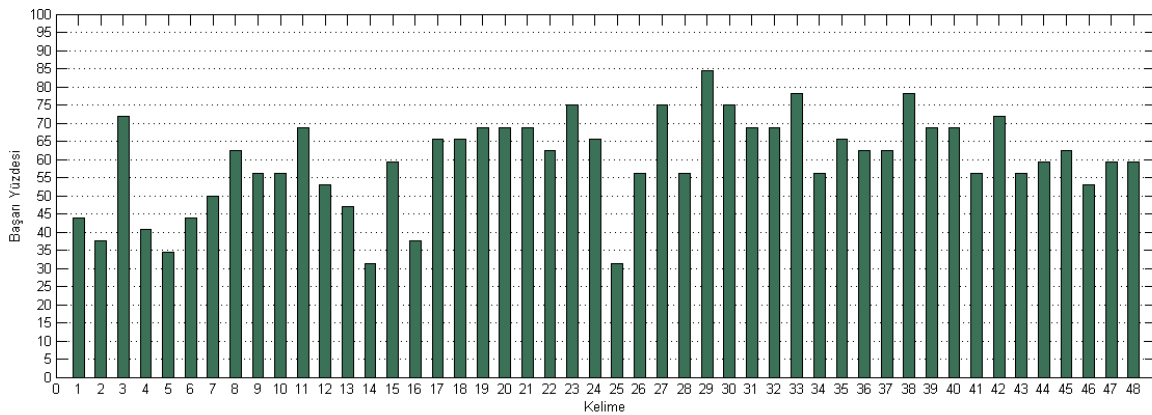
Burada 16 kHz örnekleme frekansına sahip sinyaller, yeniden örnekleme ile 8 kHz örnekleme frekansına düşürülmektedir. Çünkü MELP kodlama 8 kHz örnekleme frekansına göre uyarlanmış bir yöntemdir. Daha sonra veri tabanındaki tüm veriler MELP kodlama algoritması ile kodlanmış ve sonrasında kod açılması ile yeni veri tabanı oluşturulmuştur. Bu veri tabanı Şekil 4.1'deki konuşmacı tanıma sistemi ile çalıştırılmış ve

%59,77'lik bir başarımlar sağlanmıştır. Kişilere göre yüzde başarımlar Şekil 4.8'de verilmiştir.



Şekil 4.8. MELP kodlanıp kod açılmış konuşma veritabanıyla MFCC öznitelikleri kullanarak yapılan sınıflandırmada kişiler için yüzde başarımları

Yapılan sınıflandırmada en yüksek başarımlar oranına sahip 1. kişinin seslendirdiği kelimelerin % 76,04'ü sistem tarafından doğru kişiye eşleştirildiği görülmektedir. Sınıflandırmada en düşük tanınma oranına sahip 16. Kişinin başarımlar oranı % 31,25 olmuştur. Veri tabanı indisi 29 olan 'Bidon' kelimesinin doğru kişiye eşleşme yüzdesi %84,38 ile en yüksek orana sahiptir. 14. kelime olan "Kartopu" ve 25. kelime olan "Tam" kelimesi ise %31,25 ile doğru kişiye eşleşmede en az yüzdeye sahiptirler. Sistemin kelimelere göre yüzde başarımlar tablosu Şekil 4.9'da verilmiştir. MELP kodlamaya maruz kalmış veri tabanının MFCC öznitelikleri kullanılarak oluşturulan sistemde toplam başarımlar oranı %59,77 olmuştur.



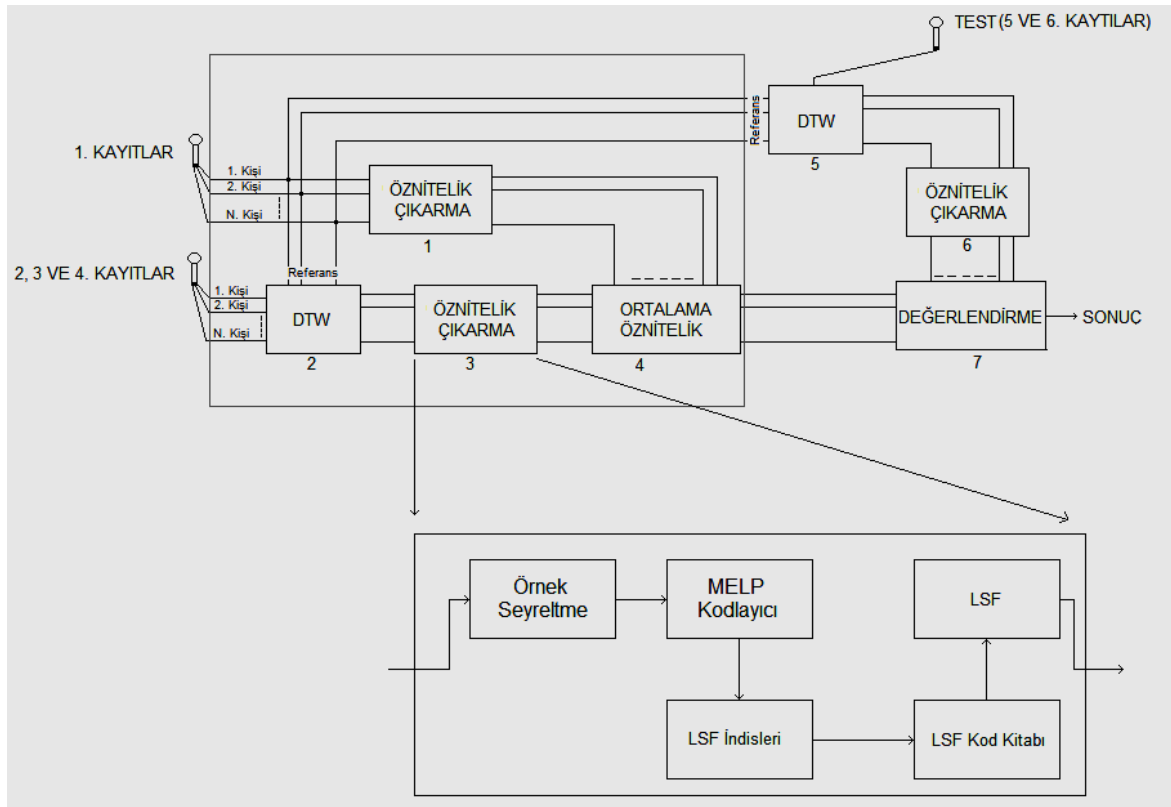
Şekil 4.9. MELP kodlanıp kod açılmış konuşma veritabanıyla MFCC öznitelikleri kullanarak yapılan sınıflandırmada kelimeler için yüzde başarımları

Bölüm 4.1’de kodlamaya maruz bırakılmamış veri tabanı ile yapılan sınıflandırma başarımı %95,51 iken, MELP kodlamaya maruz kalmış veri tabanının yine aynı sistem üzerindeki başarımı %59,77’ye düşmüştür. Kodlama işleminden kaynaklı bu başarı düşüşünü azaltmak adına Bölüm 4.3’de yeni bir yöntem önerilmiştir.

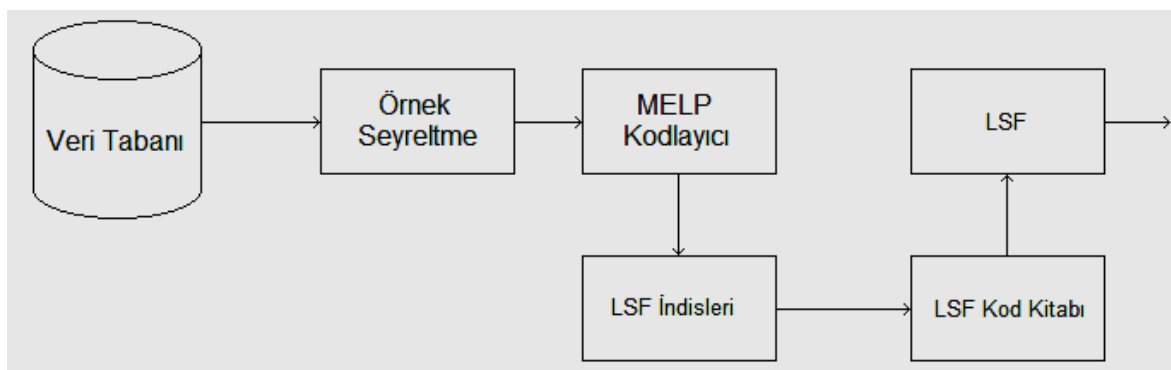
4.3. MELP Öznitelikleri Kullanılarak Sınıflandırma Başarımı

Bölüm 4.1 ve 4.2’de yapılan testler sonucunda kodlanmaya maruz bırakılmış verilerle yapılan konuşmacı tanıma işleminde, kodlanmamış sinyallerle yapılan konuşmacı tanıma sistemine nazaran çok daha düşük başarı oranı gözlemlenmiştir. Hâlbuki iki sistem arasında başka herhangi bir değişiklik yoktur. Buna rağmen başarı düşümü % 35,74 olmuştur. Bu başarı düşümünü azaltmak için kodlama işleminden elde edilen verilerin öznitelik olarak kullanılması yolu seçilmiştir. Bu sayede kodlanmış olan konuşma sinyali tekrar kod açılmasına gerek kalmaksızın sınıflandırma işlemine tabi tutulabilir. Kodlamanın bir amacının da veri boyutunu azaltıp iletim ortamında daha az veri ile daha çok bilgi iletiminin sağlanması olduğundan iletim kanalından elde edilen kodlanmış veri kullanılarak, kod açılma işlemine gerek kalmaksızın konuşmacı tanıma uygulamasının yapılabilmesi büyük bir kazanımdır. Böylelikle hem kod açma işlemi, hem de başka bir öznitelik çıkarım yöntemi kullanılmasına gerek kalmamış olur ve işlem yükü açısından da büyük bir kazanım elde edilmiş olur. Şekil 4.1’deki sistemde öznitelik çıkarma bloklarında MELP kodlama işlemi uygulanmış ve MELP kodlamadan elde edilen kodlanmış LSF bilgileri öznitelik olarak kullanılmıştır (Şekil 4.10).

Melp kodlamanın öznitelik çıkarımı amacıyla kullanıldığı konuşmacı tanıma sisteminde kullanılan veri tabanı dönüşüme uğramamış veri tabanıdır. Gerekli dönüşüm işlemi öznitelik çıkarımı bloğu içerisinde yapılmıştır. 16 kHz örnekleme frekansına sahip konuşma sinyali bölütü öznitelik bloğuna giriş yapıp, blok çıkışında o bölüte ait LSF indeksleri blok çıkışına verilmiştir. Yapılan işlemler sırasıyla örnek seyretme, MELP kodlama, kod kitabından LSF indislerinin gösterdiği vektörleri toplamak ve blok çıkışına 10 birimlik kodlanmış LSF vektörünün verilmesi olarak sıralanabilir (Şekil 4.11).



Şekil 4.10. Konuşmacı tanıma sisteminde MELP kodlamanın öznitelik çıkarımında kullanılması



Şekil 4.11. MELP kodlamanın öznitelik çıkarımında kullanılması

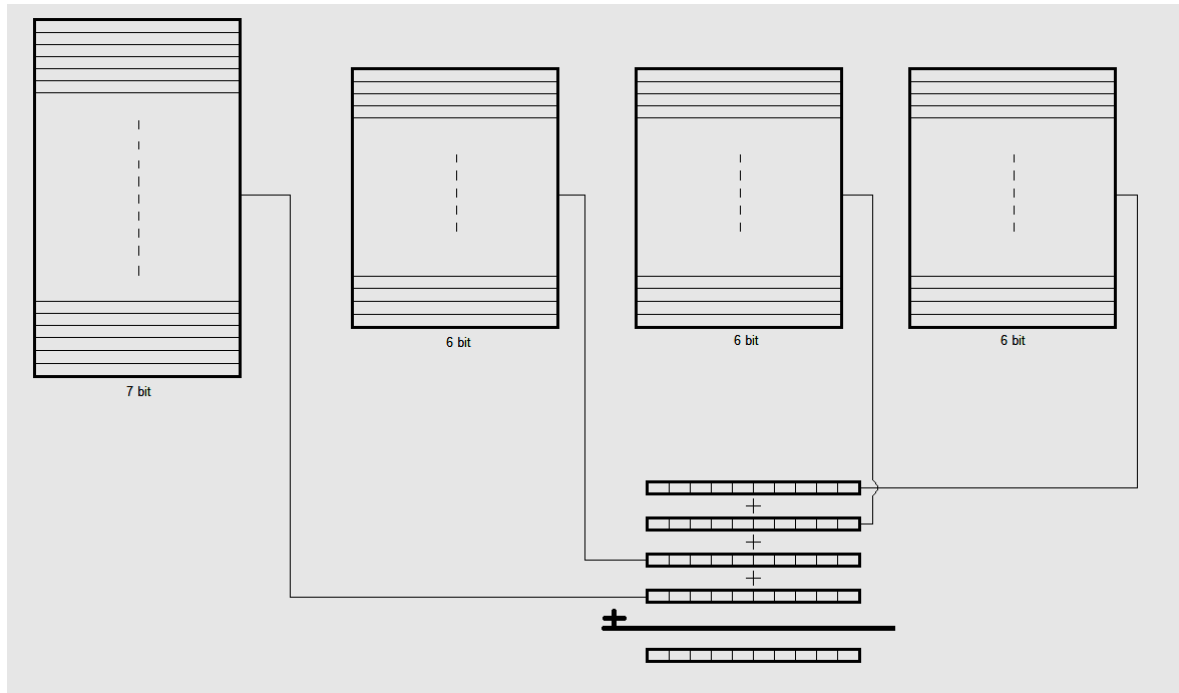
Örnek seyreltme işlemi konuşma sinyalinin MELP kodlayıcının çalıştığı 8 kHz örnekleme frekansına çekilmesi amacıyla yapılmaktadır. Sonrasında MELP algoritması ile kodlanmaktadır. Kodlanmış sinyalde o konuşma sinyali ile ilgili bazı bilgiler mevcuttur. Bunlardan biri de o konuşma bölütünün içerdiği çizgisel spektral frekanslardır (LSF). MELP kodlamada LSF'lerin tanımlı olduğu bir kod kitabı mevcuttur. MELP kodlama işleminde 10. dereceden LPC analizi yapıldığından kod kitabındaki her bir vektörün boyutu 10 birimdir. Bu kod kitabı çok katmanlı vektör nicemleyici şeklinde tasarlanmış

olup 4 katmanı bulunmaktadır. Bu dört katmanda sırasıyla 128, 64, 64 ve 64 adet vektör bulunmaktadır. Şekil 4.12’de çok katmanlı vektör nicemleyicinin her bir katmanından ikişer örnek vektör gösterilmektedir.

Katman	İndis	(1) Hz	(2) Hz	(3) Hz	(4) Hz	(5) Hz	(6) Hz	(7) Hz	(8) Hz	(9) Hz	(10) Hz
1	0	355.24	492.66	635.39	980.35	1837.26	2145.00	2413.63	2740.71	3093.75	3368.61
1	1	484.73	640.39	823.55	1338.89	1880.24	2096.15	2435.88	2765.30	3111.36	3418.14
2	0	-1.08	-11.58	2.31	-14.00	-69.46	37.77	-72.56	-159.22	-339.82	-152.12
2	1	-49.65	-46.25	-78.10	-42.88	87.14	6.81	-84.40	-169.55	114.69	84.98
3	0	32.98	12.95	-77.52	3.46	21.12	-18.52	8.59	-32.46	79.44	57.10
3	1	20.02	21.87	-34.91	79.79	42.79	37.29	53.15	-12.29	-122.97	-61.57
4	0	13.86	-19.74	20.32	-1.37	-24.03	55.40	-33.43	-7.16	-76.74	-66.68
4	1	-22.29	-34.19	20.70	-14.17	28.37	-25.98	-43.56	15.94	46.56	34.64

Şekil 4.12. Çok katmanlı MELP kod kitabından alınmış örnek vektörler

Kodlama sonucunda her bir katmandan bir indis belirlenir ve indislerin işaret ettiği vektörlerin toplamı nicemlenmiş LSF vektörünü, $\hat{f}$, oluşturur. Şekil 4.13 bu işlemi göstermektedir.



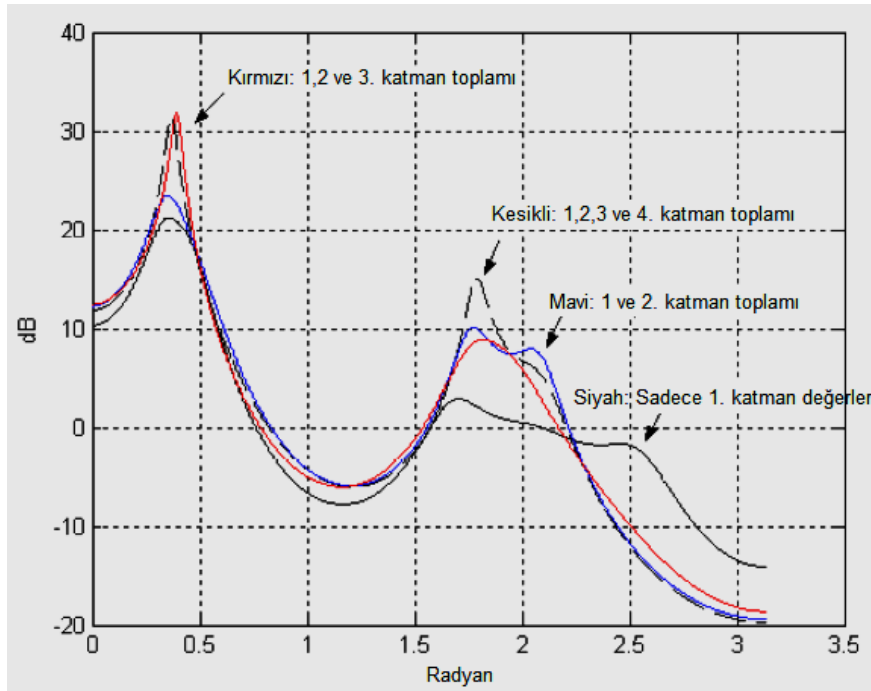
Şekil 4.13. Çok katmanlı MELP kod kitabından nicemlenmiş LSF vektörünün çıkarımı

Kodlama sırasında çok katmanlı kod kitabının indisleri, nicemlenmemiş LSF vektörü ile ağırlıklandırılmış (Eş.4.2) Öklid mesafesinin karesinin, d^2 , minimum olması kuralıyla seçilir (Eş.4.1) [21].

$$d^2(f, \hat{f}) = \sum_{i=1}^{10} w_i (f_i - \hat{f}_i)^2 \quad (4.1)$$

$$w_i = \begin{cases} P(f_i)^{0.3}, & 1 \leq i \leq 8 \\ 0.64P(f_i)^{0.3}, & i = 9 \\ 0.16P(f_i)^{0.3}, & i \leq 10 \end{cases} \quad (4.2)$$

Çok katmanlı kod kitabındaki ilk iki katmanın toplamı LSF değerlerinin kabataslak değerini oluştururken, 3. ve 4. katmanlar istenilen değere yaklaşma adına ince ayarlama maksadı ile kullanılmaktadır. Şekil 4.14'te çok katmanlı kod kitabındaki LSF vektörlerinin sadece 1. katmandaki değeri ve 2., 3. ve 4. katmandaki verilerin de eklenmesi durumunda süzgeç cevabının değişimi gösterilmiştir.



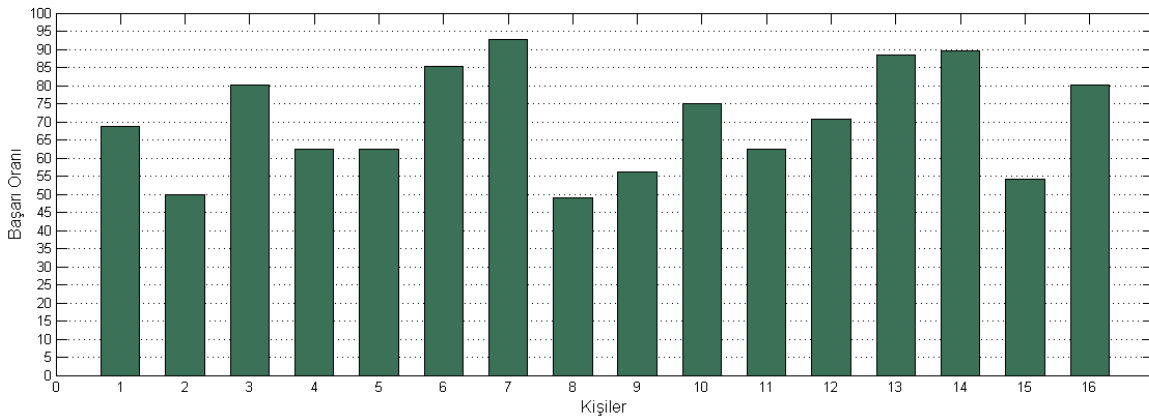
Şekil 4.14. Çok katmanlı kod kitabının katmanlarındaki değerlerin süzgeç cevabına katkısı

Çok katmanlı kod kitabının her bir katmanındaki eleman sayısı hesaba katıldığında her bir bölüt için $128 \times 64 \times 64 \times 64 = 33\,554\,432$ farklı nicemlenmiş LSF değeri oluşma olasılığı vardır. Anlamlı bir sözcüğün en az bir saniye olduğu varsayılırsa ve MELP kodlamada her bir bölütün 22.5 ms olduğu bilindiğine göre, bu sözcüğün minimum 44 bölüte sahip olacağı düşünülebilir. Bu da bir saniyelik bir kelime için $(33\,554\,432)^{44}$ farklı

kombinasyonun oluşabileceğini göstermektedir. Bu durum, nicemlenmiş LSF bilgisinin, konuşmacı tanıma uygulamalarında kullanılabilecek yeterli çeşitliliği barındırdığı manasına gelir.

MELP kodlamadan elde edilen nicemlenmiş LSF değerlerinin öznitelik olarak kullanıldığı durumda yapılan sınıflandırmada %70,51 oranında test kaydının doğru kişiye eşlendiği görülmüştür. Bölüm 4.2’de anlatıldığı şekilde kodlanıp kod açılan sinyallere MFCC öznitelik çıkarımı yapıldığı durumda sınıflandırma başarısı %59,77 idi. Öznitelik olarak MELP kodlamadaki nicemlenmiş LSF kullanımı durumunda bu oranda %10,74’lük bir iyileştirme sağlanmıştır. Bu yöntemin bir başka iyi yönü de, MELP kodlamaya maruz kalmış konuşma sinyalinin, kod açılmasına gerek kalmadan kime ait olduğunun kabul edilebilir bir başarıyla belirlenebilmesidir.

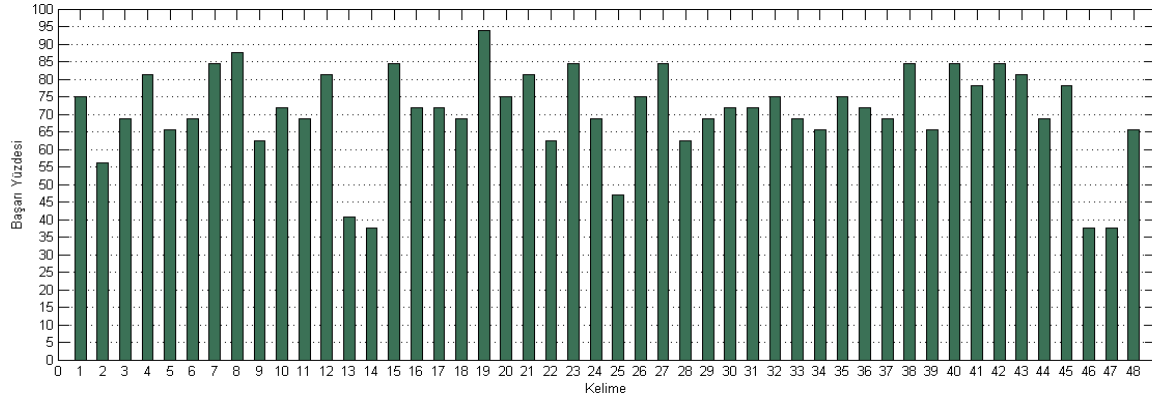
Şekil 4.15’te nicemlenmiş LSF özniteliği kullanılarak sınıflandırmada kişilere göre doğruluk yüzdesi verilmiştir. 7. Kişinin seslendirmiş olduğu kelimelerin sistem tarafından %92,71 doğruluk oranı ile kendine eşlendiği görülmüştür. %48,96 doğruluk oranı ile sınıflandırılan 8. kişi sistem tarafından en düşük başarımın sağlandığı kişi olmuştur.



Şekil 4.15. Öznitelik olarak MELP kodlamadan elde edilen nicemlenmiş LSF değerlerinin kullanılmasıyla yapılan sınıflandırmada kişiler için yüzde başarımı

Şekil 4.16’da nicemlenmiş LSF özniteliği kullanılarak sınıflandırmada kelimelere göre doğruluk yüzdesi mevcuttur. Başarı oranları incelendiğinde en yüksek oranda kime ait olduğu belirlenen kelime, veri tabanında 19. kelime olan “Tuzak” kelimesidir. “Tuzak” kelimesinin başarı oranı %93,75 olmuştur. Veri tabanında 14, 46 ve 47. indise sahip kelimeler %37,5 ile en düşük sınıflandırma başarısının gözlemlendiği kelimeler olmuşlardır. Bu kelimeler sırasıyla “Kartopu”, “Pırasa” ve “Pilav” kelimeleridir. Başarımı %50’nin

altında kalan “Kartopu”, “Pırasa”, “Pilav”, “Kapak” ve “Tam” kelimeleri yok sayılıp sistem başarımı ölçüldüğüne %70,51’den %74,06’ya çıktığı gözlenmektedir. Bu da anahtar kelime ile çalışan sistemlerde doğru anahtar kelime seçiminin sistem performansına etkisini açık bir şekilde sergilemektedir. Bu amaçla Bölüm 5’de doğru anahtar kelime seçiminin sistem performansına etkisi incelenecektir.



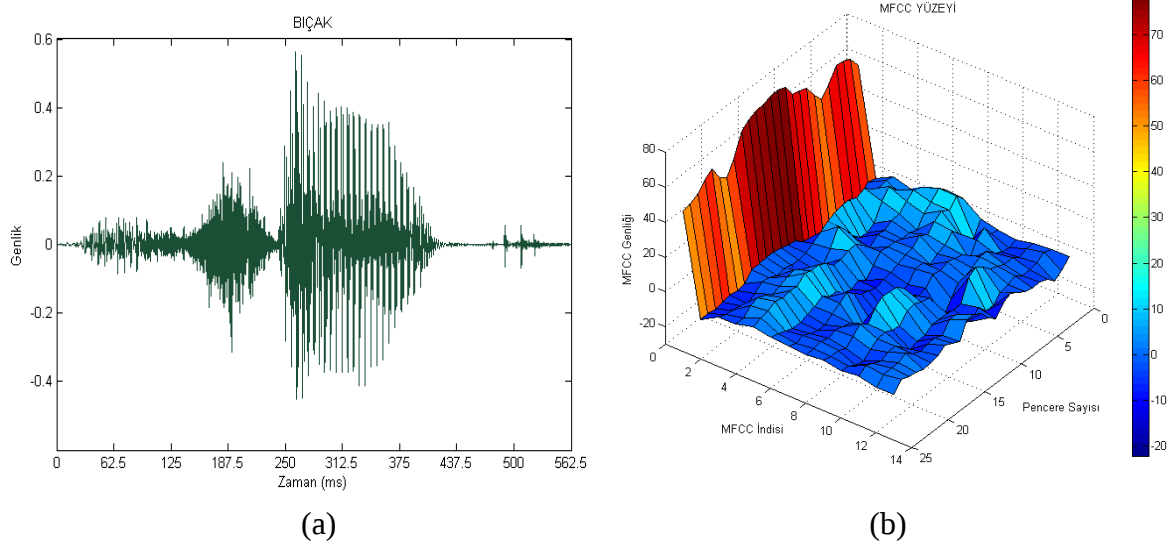
Şekil 4.16. Öznitelik olarak MELP kodlamadan elde edilen nicemlenmiş LSF değerlerinin kullanılmasıyla yapılan sınıflandırmada kelimeler için yüzde başarımı

5. MFCC ÖZNİTELİKLERİNE DAYALI KONUŞMACI TANIMADA ANAHTAR KELİME SEÇİMİ

Konuşmacı tanıma sistemlerinin güvenlik uygulamalarında çok yaygın bir biçimde kullanılamamasının nedeni olarak, insan sesinin her an ve her koşulda aynı olmaması gösterilebilir. Hastalık, yaşlanma, sigara ve alkol kullanımı ile insan sesi zamanla değişiklik gösterilebilir. İnsan kulağının duyması temel alınarak öznitelik çıkarımı yapan MFCC algoritması ile çok yüksek bir başarımlı oranı yakalamasına rağmen %100 doğruluğa çıkılamamıştır [1,25-26,29-32]. MFCC algoritmasının yanında birçok öznitelik çıkarma algoritmaları geliştirmiş, değişik sınıflandırma teknikleri uygulanmış, birçok hibrit sistem kurulmuş ancak istenilen ve her durumda %100 doğrulukla çalışan bir sistem, konuşmacı tanıma uygulamaları için henüz geliştirilememiştir. Bu amaçla özellikle anahtar kelime kullanılan kapalı küme bir konuşmacı tanıma sistemi için kullanılacak anahtar kelimenin seçimi, sistem performansı için çok önemli rol oynar. Eğer öznitelikselsel olarak kişiden kişiye değişimi yüksek, kişinin her söyleyişinin birbirine çok yakın olduğu kelimeler bu tarz sistemler için anahtar kelime olarak bulunabilirse, güvenilirlikleri son derece yüksek sistemler kurulabilir ve gerçek hayatta da kullanım imkânı olabilir. Burada, kelimelerin fonetik yapısı ve içerdiği harflerin özellikleri kritik rol oynar. Dolayısıyla uygun anahtar kelimelerin seçimi her dil için ayrı çalışma yapılmasını gerektirebilir. Bu çalışmada Türkçe kurulacak bir konuşmacı tanıma sisteminde kullanılan kelime seçiminin sistem performansına etkisi incelenmiştir. Bu amaçla 48 ayrı Türkçe kelime 16 farklı kişiye farklı zamanlarda altışar defa okutturulmuştur. İlk dört kayıt sistemin eğitimi için kullanılmıştır. Öznitelikler MFCC ile çıkartılmış, sınıflandırma Öklid mesafesi kullanılarak yapılmıştır. Bölüm 5.1 ve 5.2’de sırasıyla başarılı ve nispeten başarısı düşük sözcükler fonetik olarak incelenecektir.

5.1. Sınıflandırmada %100 Başarı Sağlayan Kelimelerin Fonetik İncelenmesi

Bıçak (İndis 28 – Tablo 3.2)



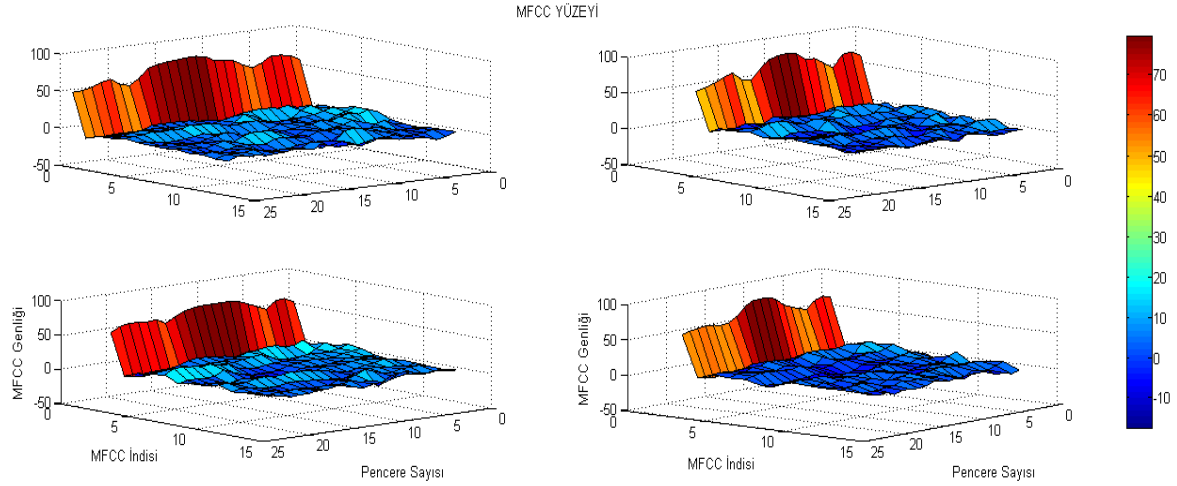
Şekil 5.1. “Bıçak” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Bıçak sözcüğü bıç-kı sözcüğündeki gibi 'bıç-'eylem kökü '-ak ve -kı' eylemden ad yapma eklerinden türemiştir. Türemiş kelimedir; ancak Türk eğitim müfredatına göre kökü unutulmuş kelimeler basit olarak öğretilip kabul edilmektedir. Eski Türkçe’de bıç- olan fiil içindeki /ç/ sesinin inceltici etkisi ile bıç- olmuştur (Ergin, 1972). Ancak günümüz Türkiye Türkçesinde bu kökün eski şeklinden türemiş bıçkı ve bıçak kelimeleri vardır.

Bıçak kelimesinde küçük ve büyük ünlü uyumları bulunmaktadır. /b/ sesi çift dudak ünsüzü olduğu için ağızdan kolay çıkmaktadır. /ı/ sesi ise ağızdan hiçbir engele uğramadan çıkabilmektedir. /ç/ sesi /t/ ve /ş/ seslerinin birleşiminden oluşan ve fonetik bağlamda sesletimi daha zor seslerden bir tanesidir. Ancak burada /ç/ sesinden önce ve sonra gelen sesler söyleyişini kolaylaştırmıştır. Kelime sonundaki /k/ sesi ise patlamalı bir sestir [33-36].

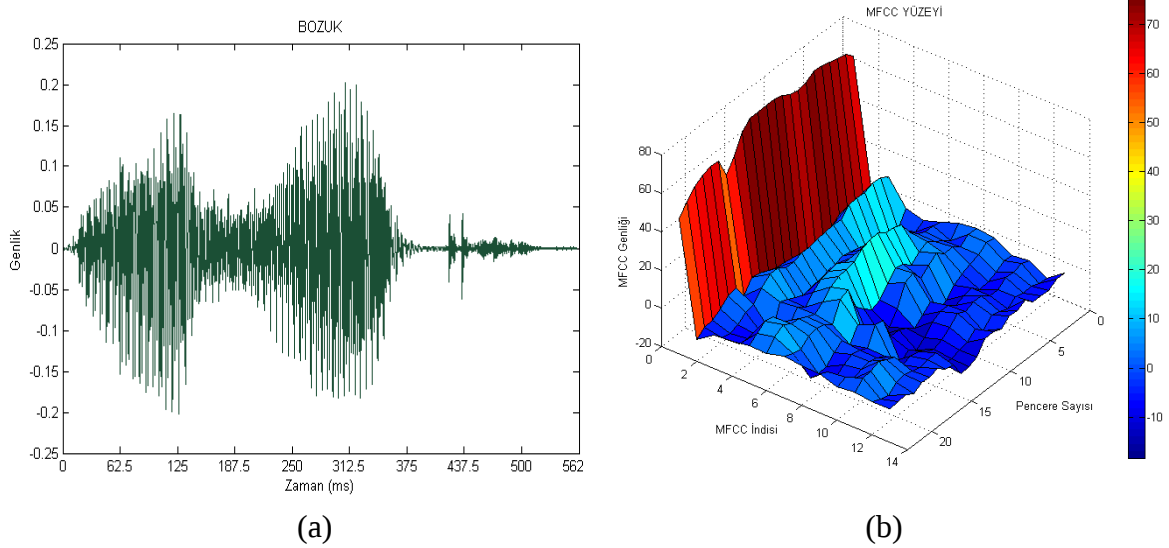
Bıçak sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.1’de verilmiştir. MFCC yüzeyleri her bir pencere için bulunan MFCC değerlerinin bir matrise yerleştirilmesiyle elde edilmiştir. Geliştirilen konuşmacı tanıma sisteminde değerlendirilen yüzeyler olduğu için, %100 başarı ile tanınan sözcüklerin MFCC yüzeylerine burada yer verilmiştir. Örnek gösterim açısından “bıçak”

sözcüğünün farklı konuşmacılardan elde edilen MFCC yüzeyleri Şekil 5.2’de verilmiştir. MFCC yüzeylerinin genel yapısının aynı olmasına rağmen, konuşmacıya bağlı olarak farklılıklar gösterdiği şekilde görülmektedir.



Şekil 5.2. Veri tabanındaki 4 kişiye ait “Bıçak” sözcüğünün MFCC yüzeyi

Bozuk (İndis 20 – Tablo 3.2)



Şekil 5.3. “Bozuk” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

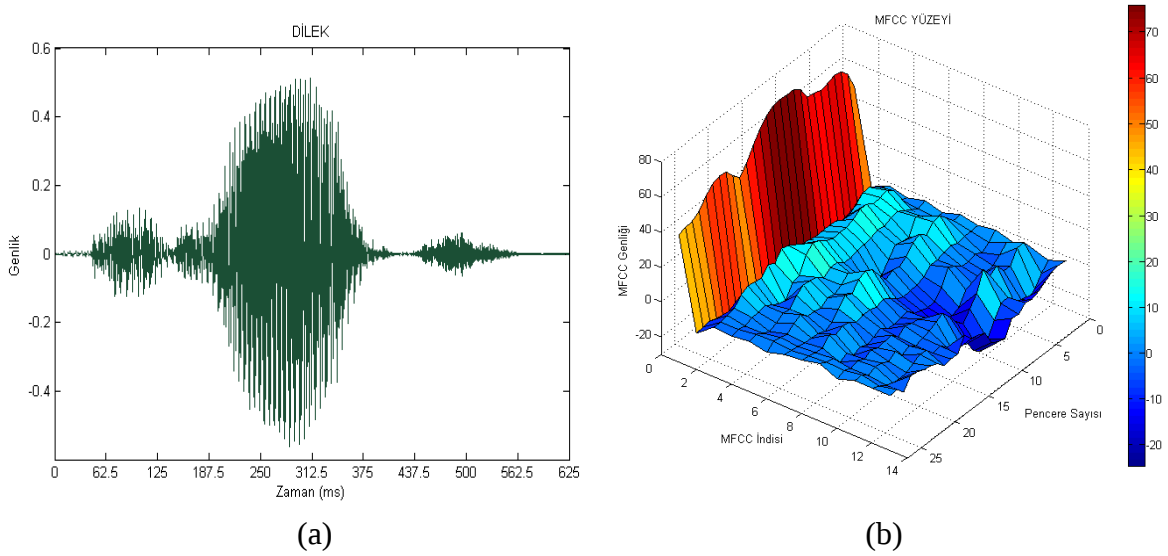
Bozuk sözcüğü Türkçe, ‘boz-’ eylem kökü ve -uk eylemden ad yapma eklerinden birini almış olduğu için türemiş bir sözcüktür.

Bozuk kelimesi küçük ve büyük ünlü uyumlarına uymaktadır. /b/ sesi çift dudak ünsüzü olduğu için ağızdan kolay çıkmaktadır. Daha sonra incelemek olan “tuzak” kelimesinin

hece yapısı ile “ bozuk” kelimesinin yapıları benzerlik göstermektedir. /z/ sesi ötümlü ve sürtünmeli /k/ sesi ise patlamalı bir sestir [33-36].

Bozuk sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.3'te verilmiştir.

Dilek (İndis 31 – Tablo 3.2)

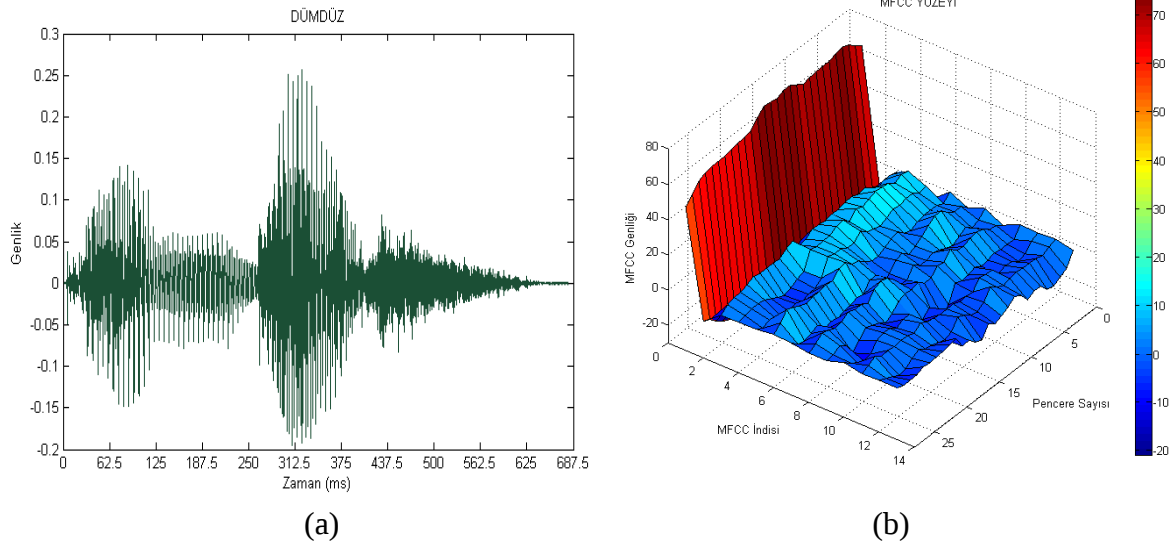


Şekil 5.4. “Dilek” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Dilek sözcüğü Türkçe kökenlidir. Dil sözcüğüne getirilen -e (isimden fiil yapım eki) ve -k (fiilden isim yapım eki) ekleri getirilerek oluşmuştur. Ünlü uyumları bakımından Türkçenin ses yapısına uygundur. Kelime başındaki /d/ sesi ötümlü ve patlamalı bir sestir. /l/ sesi ağızdan çıkış yeri bakımından dil ucu ön damak ve ötümlü ünsüzdür. Kelime sonundaki /k/ sesi ise patlamalı bir sestir [33-36].

Dilek sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.4'te verilmiştir.

Dümdüz (İndis 6 – Tablo 3.2)

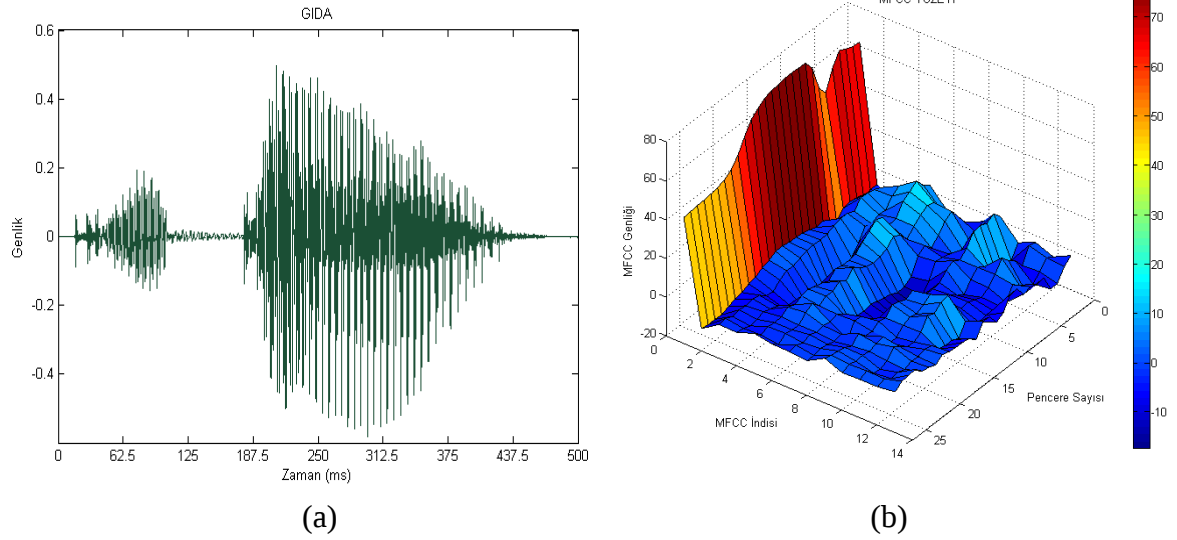


Şekil 5.5. “Dümdüz” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Dümdüz sözcüğü, pekiştirmeli bir sözcüktür. Pekiştirmeli sözcük oluşturma'nın bir biçimi, sözcüğün ilk hecesine "p, r, s, m" pekiştirme harflerinden birisinin getirilmesidir. Dümdüz kelimesi de bu şekilde oluşmuştur. Ünlü uyumları bakımından Türkçenin ses yapısına uygundur. Kelime başındaki /d/ sesi ötümlü ve patlamalı bir sestir. İkiz ses ve heceler peş peşe kullanıldığı için sesletimi kolaylaştırmıştır [33-36].

Dümdüz sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.5'te verilmiştir. Sistemdeki en yüksek başarıyı gösteren sözcük, Öklid mesafesi değerlendirildiğinde, “dümdüz”dür.

Gıda (İndis 42 – Tablo 3.2)

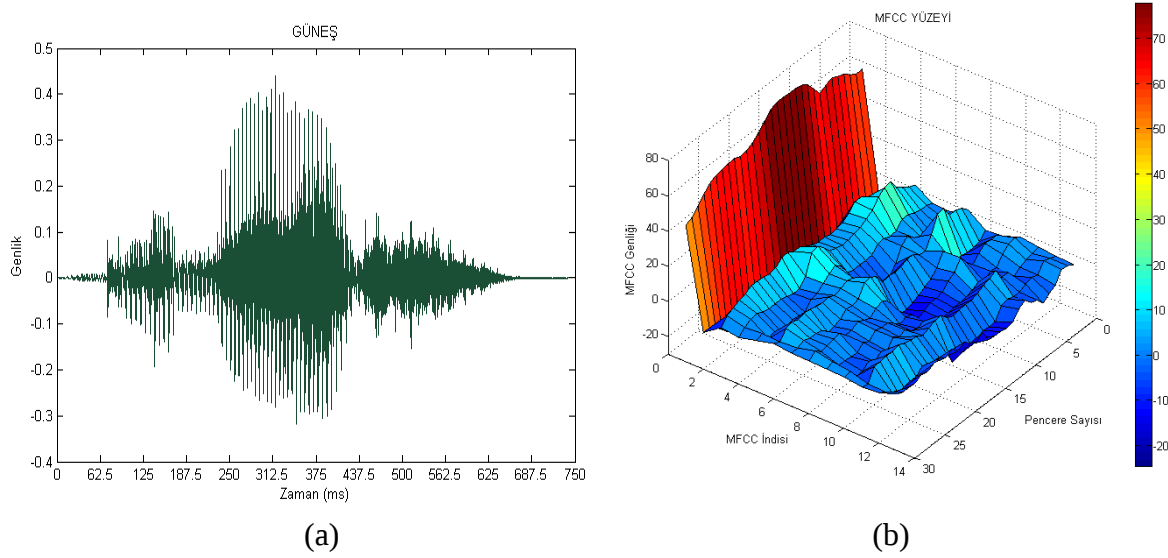


Şekil 5.6. “Gıda” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Gıda sözcüğü Türkçeye Arapçadan geçmiş sözcüklerden bir tanesidir. “Gıda” kelimesinin hece dizilimi Türkçedeki hece dizilimine uygundur ve kelime içerisinde ünlü uyumu bulunmaktadır. /g/ ve /d/ sesleri ötümlü, art damaksıl ve patlamalı seslerdir. Ancak /d/ sesi aynı zamanda genizsil bir sestir. Aralarındaki bu farktan dolayı sesletimleri kolaylaştırmıştır [33-36].

Gıda sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.6’da verilmiştir.

Güneş (İndis 43 – Tablo 3.2)



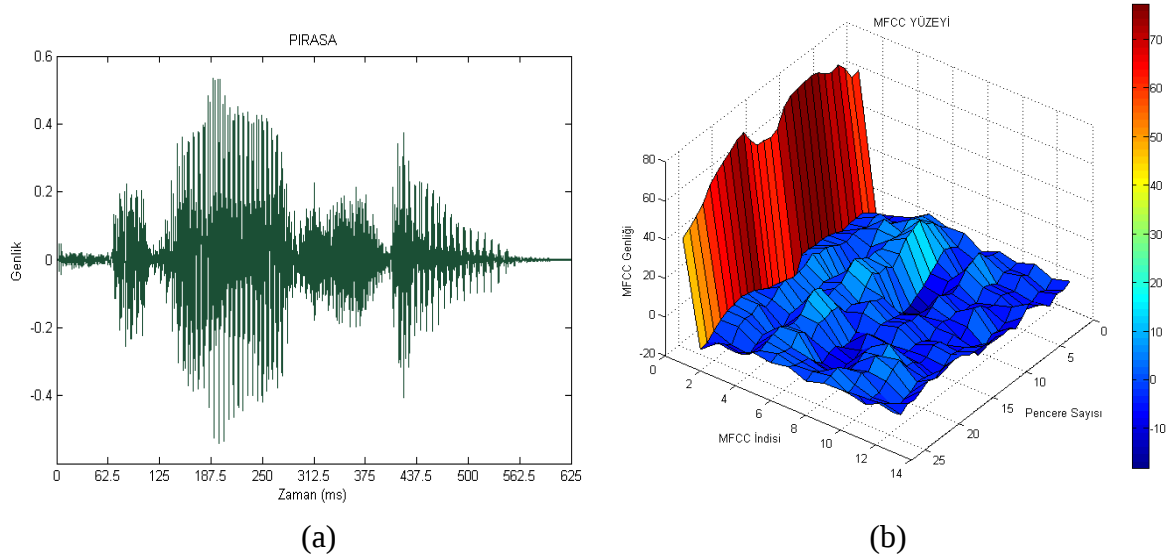
Şekil 5.7. “Güneş” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Zeynep Korkmaz Türkiye Türkçesi Grameri adlı eserinde gündüz kelimesini türemiş bir kelime olarak kabul etmektedir. Gün- isim kökü yine isimden isim yapan bir ek alarak türemiştir. Bu bakımdan güneş kelimesinin de –gün isim kökünden türemiş olduğunu söyleyebiliriz.

Dilin konumuna ve dudakların durumuna göre yuvarlak ünlüden sonra düz geniş ünlü gelir. Güneş kelimesinde de olduğu gibi bu durum Türkçedeki ses uyum kuralına uygundur. /g/ sesi ötümlü, patlamalı ve art damaksıl bir sestir. /n/ sesi ise genizsil ve patlamalı bir sestir. İkisi de patlamalı ve ağız arkası sesler olduğu için sesletimleri kolaydır. /ş/ sesi sürtünmeli bir ses olduğu için –n sesinden sonra sesletimi bitiren bir özelliğe sahiptir [33-36].

“Güneş” sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.7’de verilmiştir.

Pırasa (İndis 46 – Tablo 3.2)



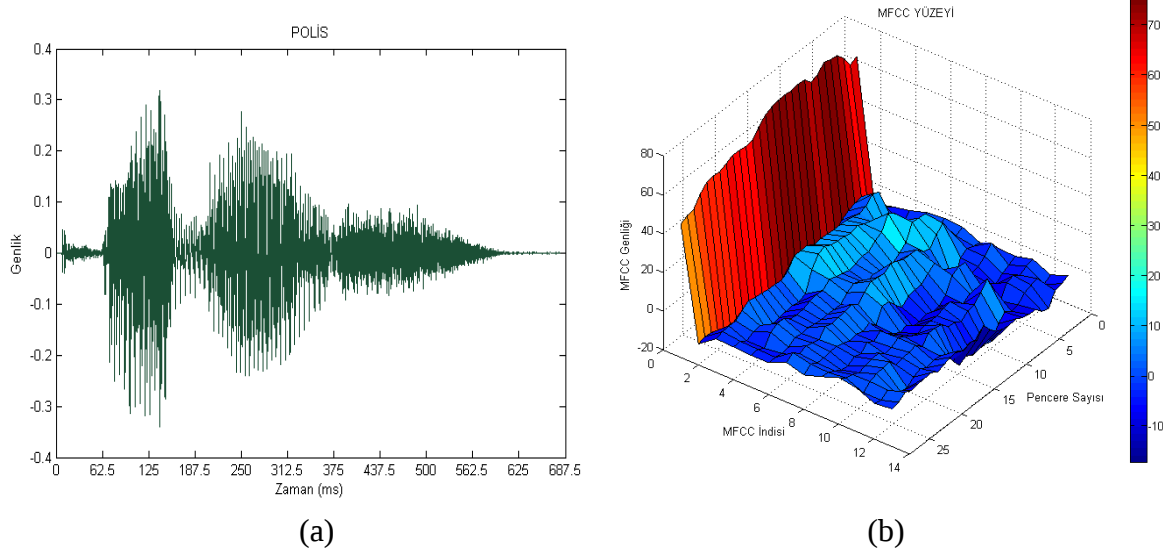
Şekil 5.8. “Pırasa” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Pırasa kelimesinin aslı, Farsça “pür-hassa” dan gelmektedir. Türkçenin ses özelliklerine uyarlanarak pırasaya evrilmiştir.

“Pırasa” kelimesinde ünlü uyumu bulunmaktadır. /ı/ düz ünlüsünden sonra /a/ düz ünlüsü gelmiştir. Aralarındaki ünlü uyumunun yanısıra bu iki ses (-ı,-a) ağız içerisinde hiçbir engele uğramadan çıkar. Türkçede bir hece ünlü ses ile bitiyorsa ötümsüz ünsüzle başlayan başka bir heceyle eklenir. Burada ilk heceden sonra gelen /r/ sesi ötümlü olmasına rağmen çarpmalı bir ses olduğu ve sonrasında /s/ ötümsüz sesi ile başlayan bir heceye eklendiği için söylenişi zorlaştırmamıştır. Bir diğer özelliği de /r/ ve /s/ seslerinin sürtünmeli sesler olmasıdır. Bu iki sesin ardıl olarak gelmesi ve iki ses arasına giren /a/ sesinin kolay çıkmasını sağlamıştır [33-36].

“Pırasa” sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.8’de verilmiştir.

Polis (İndis 44 – Tablo 3.2)

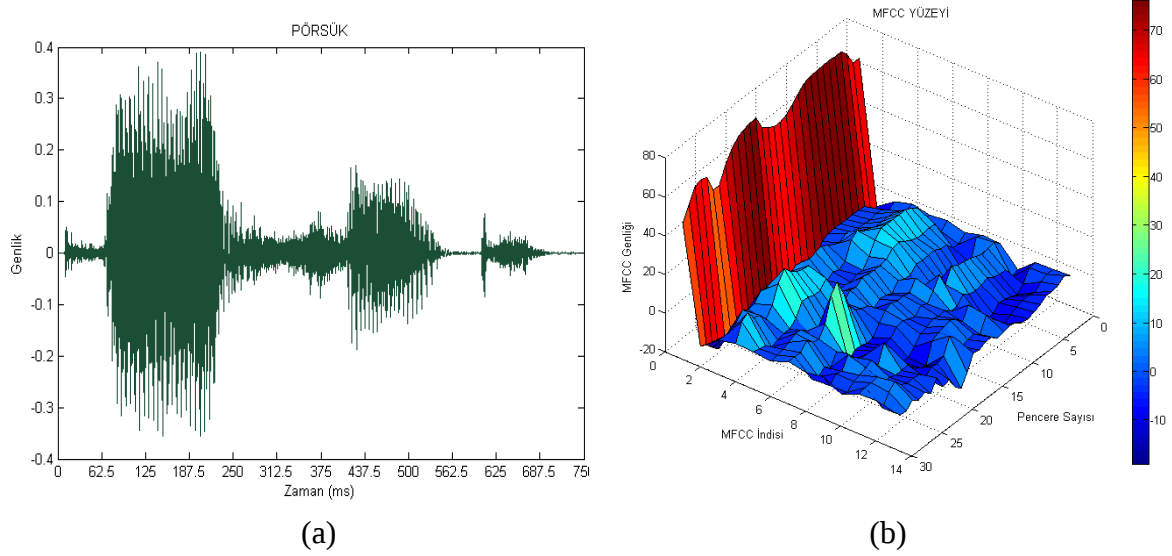


Şekil 5.9. “Polis” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Anadil konuşurlarının dillerine başka dillerden transfer olmuş kelimeleri daha zor seslendirdikleri bilinmektedir. Polis kelimesi Türkçeye Fransızcadan geçmiştir. Ancak buna rağmen bu kelimenin konuşmacılar tarafından başarılı bir şekilde söylendiği görülmektedir. Bu kelime içerisinde fonetik bağlamda sesletilmesi kolay iki sestən söz edilebilir. Bunlar; /p/ ve /o/ ‘dur. /p/ sesinin ötümsüz çift dudak ünsüzü olması ve arkasından gelen /o/ sesinin hiçbir engelle uğramaması kolay telaffuzu sağlamıştır [33-36].

“Polis” sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.9’da verilmiştir.

Pörsük (İndis 45 – Tablo 3.2)



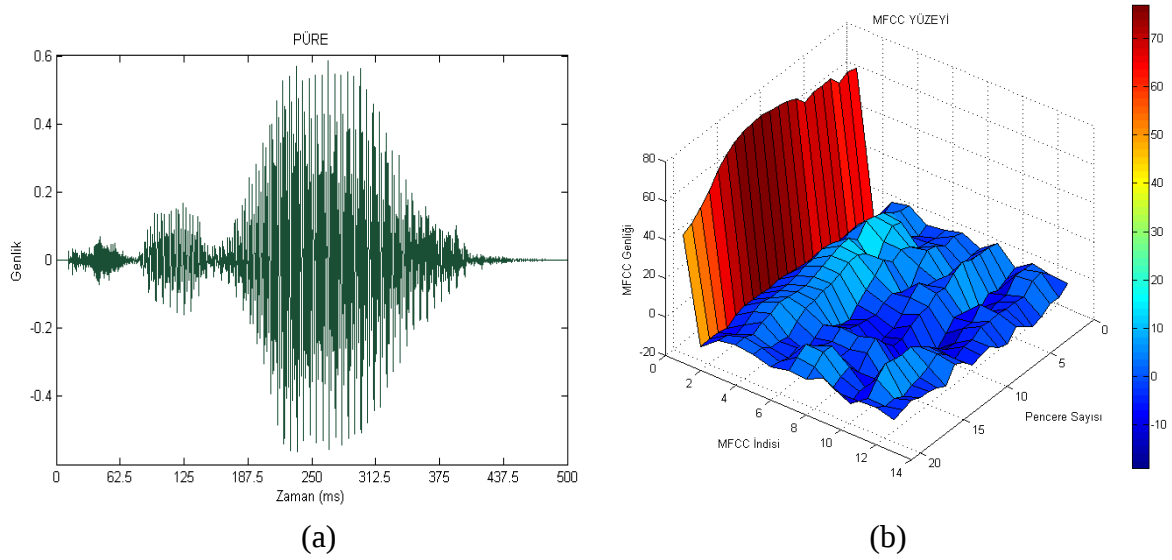
Şekil 5.10. “Pörsük” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

“Pörsük” kelimesinin yapısı pörsü-k olarak çözülebilir. –k fiilden isim yapım eki ile türetilmiş bir sıfattır.

Türkçenin ses sistemine bakıldığında ünsüz seslerin daha kolay sesletildiği görülmektedir. /p/ sesi patlamalı /r/ sesi ise çarpmalı bir sestir. Patlamalı ses sesin oluşumu sırasında engele uğrar. Fakat daha yavaş bir hızla sızarak dışarı çıkar. /r/ sesi ise hızlı bir şekilde çarparak hava akımının dil ucuna doğru kıvrılması ile oluşur. /s/ sesi sürtünücü ötümsüz bir ses iken, /k/ sesi patlamalı art damaksıl bir sestir. İlk ve ikinci hecenin çıkış yerleri açısından zıtlıkları birbirine benzeyen seslerden oluşur. /o/ ve /u/ sesleri arasında uyum vardır. Bu durum tuhaf bir biçimde söyleyiş kolaylığı sağlar [33-36].

“Pörsük” sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.10’da verilmiştir.

Püre (İndis 48 – Tablo 3.2)

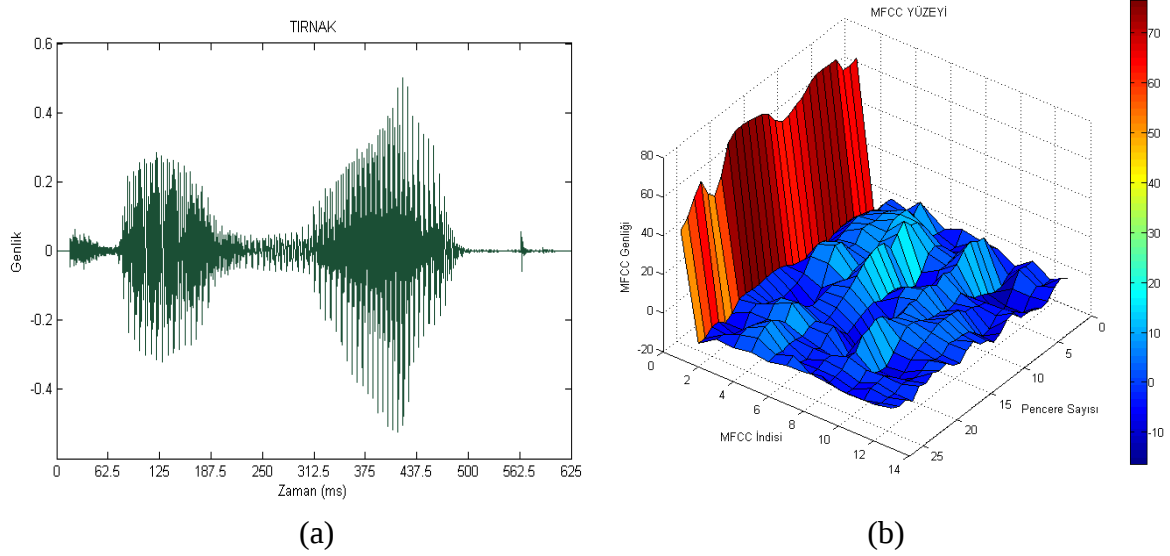


Şekil 5.11. “Püre” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Fransızca “purée” sözcüğünden alıntıdır. [37]. Ötümsüz sesler hava akımının ses tellerini gırtlak kenarına çekmesi ile oluşan ve gırtlak boşluğundan geçerek oluşturulan seslerdir. /p/ sesi ötümsüz çift dudaksıl bir sestir. /u/ sesi ise hiçbir engele uğramadan ağız içerisinde hareket eder. Püre sözcüğünün ilk hecesi bu nedenle kolayca söylenebilmektedir. /r/ sesinin ise ilk hecede sesletilme özelliği hece sonu ve ortasına göre daha kolaydır. /r/ sesi öncül sestir ve diş yuvasından çıkmaktadır. /ü/ ve /e/ sesleri arasında ise ünlü uyumu bulunmaktadır. Bu durum Türkçenin ses dizim yapısına uymaktadır. Püre kelimesi bu özellikleri bakımından kolay seslendirilmektedir [33-36].

“Püre” sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.11’de verilmiştir.

Tırnak (İndis 17 – Tablo 3.2)

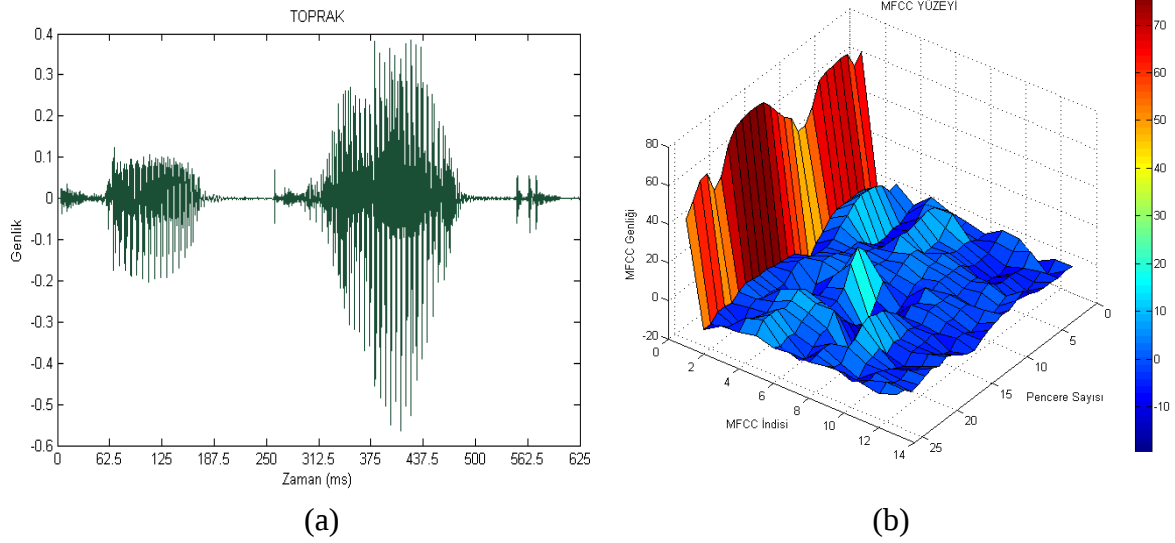


Şekil 5.12. “Tırnak” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Eski Türkçe tırna- veya tırma- fiilinden +İk sonekiyle türetilmiştir. “Tırnak” kelimesi hece dizilimi ve yapısı bakımından birazdan incelenecek olan “Toprak” kelimesi ile benzerlik gösterir. Büyük ve küçük ünlü uyumu vardır. Kelimenin ilk hecesindeki /t/ sesi ötümsüzdür ve çarpmalı bir ses olan /r/ sesini almıştır. /t/ sesi sözcüğün başında olduğu zaman ön diş sesi olması bakımından kolay çıkarılmaktadır. /r/ sesinin hece sonunda kolay söyleniyor olması ilk hecenin sesletimini kolaylaştırmıştır. Patlamalı /k/ ünsüzü kelime sonunda kullanıldığında rahat bir çıkış sağlar burada da benzer bir kolaylık oluşmuştur [33-36].

“Tırnak” sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.12’de verilmiştir.

Toprak (İndis 3 – Tablo 3.2)

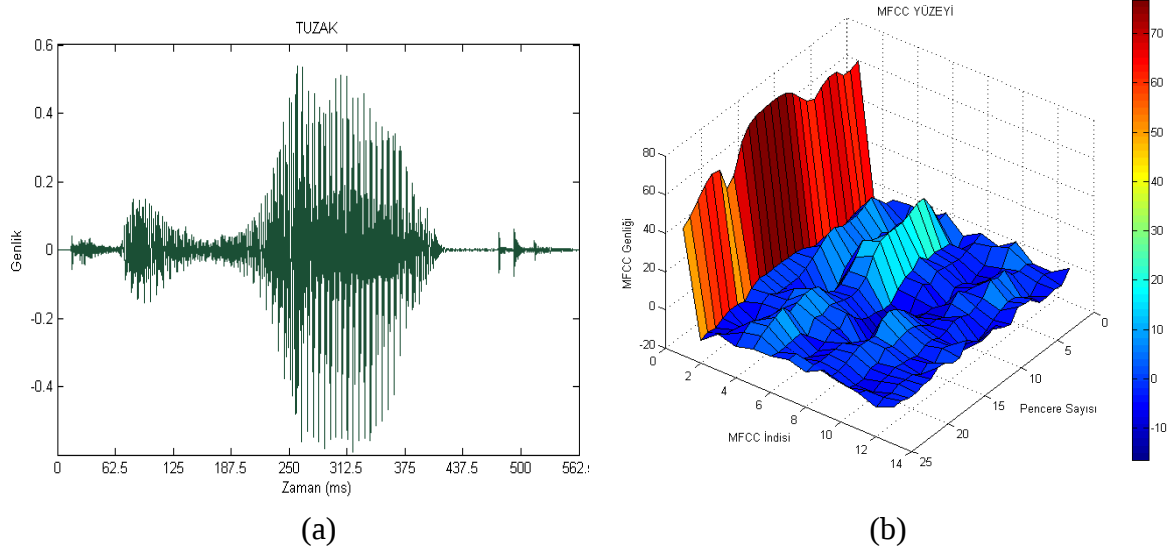


Şekil 5.13. “Toprak” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Eski Türkçe topur- fiilinden +Ik sonekiyle türetilmiştir. Toprak kelimesi içerisinde büyük ve küçük ünlü uyumları mevcuttur. /t/ sesi sözcüğün başında olduğu zaman ön diş sesi olması bakımından kolay çıkarılmaktadır. Ötümsüz /t/ ve ötümsüz /p/ seslerinin bir arada kullanılması ve bu benzerlikleri kolay çıkmalarını sağlamıştır. /p/ sesi patlamalı /r/ sesi ise çarpmalı bir sestir. Patlamalı ses sesin oluşumu sırasında engele uğrar. Fakat daha yavaş bir hızla sızarak dışarı çıkar. /r/ sesi ise hızlı bir şekilde çarparak hava akımının dil ucuna doğru kıvrılması ile oluşur. Bu iki sesin arka arkaya kullanılması ve kelime sonunda patlamalı ses olan /k/’nin kullanılması sesletime kolaylık sağlamıştır [33-36].

“Toprak” sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.13’te verilmiştir.

Tuzak (İndis 19 – Tablo 3.2)



Şekil 5.14. “Tuzak” sözcüğünün (a) zaman bölgesi ve (b) MFCC yüzeyi gösterimi

Eski Türkçe tu- "durdurmak, tutmak" sözcüğünden türetilmiş olabilir; ancak bu kesin değildir. Türeyiş biçimi açık olmamakla birlikte Eski Türkçe tur- (durmak) ve tut- (durdurmak, yakalamak) fiilleriyle bağlantılı olduğu muhakkaktır. Dönüşlü fiil yapan +Iz- ekiyle *tuz- (yakalanmak) fiili düşünülebilir; ancak bu fiil kaydedilmemiştir.

Tuzak kelimesi büyük ve küçük ünlü uyumlarına uymaktadır. /t/ sesi sözcüğün başında olduğu zaman ön diş sesi olması bakımından kolay çıkarılmıştır. Daha önce incelenmiş olan “bozuk” ve “tuzak” kelimeleri fonetik olarak birbirine benzerlik göstermektedir. /z/ sesi ötümlü ve sürtünmeli –k sesi ise patlamalı bir sestir. Tuzak ve Bozuk kelimelerinde ilk hecedeki ünlü sesler /o/ ve /u/’nun çıkış yerleri aynıdır [33-36].

“Tuzak” sözcüğünün sesletiminin veri tabanındaki bir kişiden alınmış zaman bölgesi ve MFCC yüzeyi gösterimi Şekil 5.14’te verilmiştir.

5.2. Sınıflandırmada %100 Başarı Sağlamayan Kelimelerin Fonetik Analizi

Türkçeye başka dillerden geçmiş kelimelerin sesletiminin Türkçe kelimelere nazaran daha zor olduğu bilinmektedir. Sınıflandırmada %100 başarı sağlamayan kelimeler içerisinde bu özelliklerinden dolayı bulunan kelimeler vardır. Bunlardan “gofret”, “aktör”, “garanti”, “gitar”, “ekonomik” Fransızca kökenli, “hükümet” ise Arapça kökenli örnekleridir. Bu

kelimelerin, fonetik yapıları incelendiğinde Türkçenin ses yapısına uymadıkları görülmektedir. Bu nedenle de sesletimleri nispeten daha zordur. Burada aynı zamanda /r/ sesinin kelime sonuna geldiğinde yuvarlanmasından kaynaklanan zor söylenmesine de örnek kelimeler görülmektedir. Bu kelimeler “gitar” ve “aktör” kelimeleridir. Bu duruma diğer örnekler; “pazar” , “damar” ve “tekrar” kelimeleridir. Ayrıca “hükümet” kelimesinde /h/ ve /k/ gibi boğaz seslerinin ince ünlülerle bir arada kullanılması da sesletimi zorlaştırmaktadır.

“Perşembe” kelimesi örneğinde çıkış yeri aynı olan sesler bir arada kullanıldığı için sesler birbirine karıştırılarak söylenir. Türkçe’de /g/ ve /n/ sesleri bir arada kullanıldığında söyleyişi zorlaştırmaktadır. “Geniş” ve “garanti” örneklerinde de bu durum söz konusudur. “Tam” gibi tek heceli kelimelerde sürekli sessizlerin hece sonunda kullanılması sesletimi zorlaştırır. /b/, /p/, /m/ gibi patlayan sesler hece sonunda zor, hece başında ise kolay söylenmektedir. “Dolap” ve “tam” kelimeleri bu duruma örnektir. “Dümdüz” kelimesi örneğinde ikiz hecelerin peş peşe kullanıldığında söyleyişi kolaylaştırdığı görülmüştür. Ancak “Tütün” örneğinde olduğu gibi söyleyişi kısmen zorlaştırdığı örneklerin olduğunu da söylemek mümkündür. Bu durum ikiz hecelerin peş peşe kullanıldığı kelimeler sürekli sessizle biterse rahat, süresiz sessizle biterse zor söylenir biçiminde ifade edilebilir. Türkçede kelimedeki içerisindeki seslerin çıkış yerleri aynı olduğu takdirde sesletim zorlaşır. Bu bakımdan, “buruşuk” ve “bütün” kelimeleri başarımla düşüklüğünün nedenlerinden biri de bu olabilir.

Sonuç olarak konuşmacı tanıma başarımları açısından görülmüştür ki, söylenişi kolay Türkçe kökenli sözcükler seçildiği zaman sistem daha başarılı çalışmaktadır. Bunun sebebinin kişilerin söylenişi kolay olan sözcükleri daha tutarlı seslendirmeleri ve buna bağlı olarak sistemin daha başarılı çalışması olduğu söylenebilir.

6. SONUÇ VE ÖNERİLER

Bu tez çalışması kapsamında, ayrışık sözcük kullanımıyla çalışan bir konuşmacı tanıma sistemi geliştirilmiştir. Geliştirilen konuşmacı tanıma sisteminin Türkçe sözcükler için başarımını ölçmek adına bir veri tabanı oluşturulmuştur. Oluşturulan veri tabanında 16 konuşmacının 6 tekrarla söylediği 48 farklı kelime mevcuttur. Veri tabanının, oluşturulan konuşmacı tanıma sisteminde MFCC öznitelikleri kullanıldığı durumda eğitilip, test edilmesi sonucunda sistemin başarımı %96 olduğu gözlenmiştir. Sonrasında, sistemin aynı öznitelik çıkarım metodu ile kodlamaya karşı gürbüzlüğünün testi adına, oluşturulan veri tabanı MELP kodlama ve kod açma işlemine tabi tutularak yeni veri tabanı oluşturulmuştur. Bu veri tabanı kullanılarak MFCC öznitelikleri ile yapılan konuşmacı tanıma işleminde başarım oranı %60 olmuştur. Gözlenen bu başarım düşümünü iyileştirmek için MELP kodlamanın kendisinin öznitelik olarak kullanılması yoluna gidilmiştir. MELP kodlamada kullanılan nicemlenmiş LSF değerleri kullanımı ile başarım yaklaşık %11 arttırılmış ve %71 olmuştur. Bu şekilde MELP kodlanmış konuşma sinyalinin bir daha kod açılmasına gerek kalmaksızın konuşmacı tanıma uygulaması yapılmış ve konuşma sinyaline tekrar kod açıp, kod açılmış haline öznitelik çıkarım işlemi uygulama gereği ortadan kaldırılmıştır. Böylece sadece çok katmanlı vektör nicemleyicideki 4 adet 10 birimlik vektörün toplanması işlemi ile öznitelik çıkarma işlemi tamamlanmış olmakta ve çok fazla olan işlem yükü azaltılmış olmaktadır. MELP kodlamada elde edilen kodlanmış veride, MELP'in nicemlenmiş LSF değerlerinden başka, kişiden kişiye değişimi yüksek olan öznitelikler kullanılarak desteklenmesi durumunda başarımın çok daha fazla artabileceği düşünülmektedir.

Tezin son bölümünde, MFCC öznitelik çıkarım yöntemiyle sistemin, kişinin kim olduğunu %100 doğrulukla ayırt ettiği kelimeler belirlenmiştir. Bu başarımı sağlayan 13 farklı sözcüğün fonetik yapısının bu başarımdaki olası katkısı incelenmiştir. Veri tabanındaki kelimelerin incelenmesi sonucunda, sesletimi kolay, yabancı kökenli olmayan sözcüklerde MFCC özniteliklerine göre konuşmacı tanıma başarımının, diğer sözcüklere göre daha yüksek olduğu gözlenmiştir. Bunun sebebinin sesletimi kolay sözcüklerin kişiler tarafından her okumada daha tutarlı bir şekilde seslendirilmiş olduğu düşünülmektedir. Tutarlı tekrarlarla eğitilen ve test aşamasında da benzer şekilde seslendirilen sözcüklerle sistem başarımı artmaktadır. Buradan hareketle, güvenlik uygulamalarında anahtar kelime seslendirilmesi esasıyla çalışan konuşmacı tanıma sistemlerinin tasarımı, kullanılan

öznitelik çıkarma yöntemi, kullanılan sınıflandırma algoritmalarının seçiminin yanı sıra, sistemde kullanılacak olan sözcük veri tabanındaki kelimelerin seçimi de çok önem arz ettiği görülmüştür. Bu çalışmaya ek olarak, başarıyı yüksek çıkan kelimeler ile benzer fonetik özelliklere sahip bir veri tabanı oluşturarak, sistem performansı bu veri tabanı için hesaplanabilir. Öngörülen şekilde bu veri tabanından elde edilen başarımın çok yüksek olması durumunda, güvenlik amaçlı kurulan ayırık sözcük tabanlı Türkçe konuşmacı tanıma sistemleri için özel kelime havuzu oluşturulabilir.

KAYNAKLAR

1. Tekbaş, Ö.H., Demir, M., Koçer, H. ve Yıldırım, Ü. (2013). Ses İşaretleri İçin Parametrik Olmayan Çok Boyutlu Bir Modelleme Yaklaşımı. *Savunma Bilimler Dergisi*, 12(1), 1-19.
2. Campbell, J. P., (1997). Speaker Recognition: A Tutorial. *Proc. IEEE* 85(9), 1437-1462.
3. Dunn, R.B., Quatieri, T.F., Reynolds, D.A., Campbell, J.P. (2001). Speaker Recognition from Coded Speech and the Effects of Score Normalization. Signals, Systems and Computers. *Conference Record of the Thirty-Fifth Asilomar Conference on*, 2.
4. Claes, T., Dologlou, I., Bosch, L. ve Compernelle, D. (1998). A Novel Feature Transformation for Vocal Tract Length Normalization in Automatic Speech Recognition. *IEEE Trans. On Speech and Audio Processing*, 6, 549-557.
5. Nemer, E., Goubran, R., Mahmoud, S. (2001) Robust voice activity detection using higher-order statistics in the LPC residual domain. *IEEE Trans. Speech Audio Processing*, 9(3), 217-231.
6. Atal, B. ve Schroeder, M. (2003). Predictive Coding of Speech Signals and Subjective Error Criteria. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 27(3), 247-254.
7. Atal, B. ve Rabiner, L. (1976). A Pattern Recognition Approach to Voiced unvoiced-Silence Classification with Applications to Speech Recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(3), 201-212.
8. Şentürk, Z., Yetgin, Ö.E., Salor, Ö. (2014). Özilinti Matrisi Kullanılarak Ötümlü-Ötümsüz Konuşma Sınıflandırma. *IEEE 22nd Signal Processing and Communications Applications Conference*, 1802-1805.
9. İnternet: Speech production and perception tools for real-time assessment, analysis and therapy product. Web: www.laryngograph.com adresinden 8 Mart 2015'te alınmıştır.
10. Hui, L., Dai, B., Wei, L. (2006). A Pitch Detection Algorithm Based on AMDF and ACF. *IEEE International Conference ICASSP*, 1, I.
11. Ying, G.S., Jamieson, L.H. ve Michell, C.D. (1996). A probabilistic approach to AMDF pitch detection. *Spoken Language, ICSLP 96*, 201-2204.
12. Noll, A.M. (1967). Cepstrum Pitch Determination. *Journal of the Acoustical Society of America*, 41(2), 293-309.
13. Hess, W. (1982). Pitch determination of speech signals: algorithms and devices. Automatic Speech Analysis and Recognition NATO Advanced Study Institutes Series, Springer, 88(1), 49-67.
14. Fant, G. (1960). Acoustic Theory of Speech Production. Mouton & Co, The Hague, Netherlands.

15. Rabiner, L.R. ve Schafer, R.W. (1978). Digital Processing of Speech Signals. *Signal Processing Series, Prentice Hall*.
16. Dutoit, T. (1996), An Introduction to Text-To-Speech Synthesis. *Kluwer Academic Publishers*.
17. Yang, P.F. ve Stylianou, Y. (1998). Real Time Voice Alteration Based on Linear Prediction. *Proc. of the International Conference on Spoken Language Processing, ICSLP'98*.
18. Markel, J.D. and Gray, J.A.H. (1976). Linear Prediction of Speech. *Springer-Verlag*.
19. Arslan, L.M. (1999). Speaker Transformation Algorithm Using Segmental Codebooks. *Speech Communication*, 28(3), 211-226.
20. Paliwal, K.K. (1995). Interpolation Properties of Linear Prediction Parametric Representations. *Proc. of the European Conference on Speech Communication and Technology, EUROSPEECH'95*.
21. Federal Information Processing Standards Publication (1997). *Specifications for the Analog to Digital Conversion of Voice by 2400 Bit/Second Mixed Excitation Linear Prediction*.
22. Itakura, F. (1975). Line spectrum representation of linear prediction of speech signals. *Journal of Acoustical Society of America*, 57, 35.
23. Pellom, B. L. (1998). *Enhancement, Segmentation and Synthesis of Speech with Application to Robust Speaker Recognition*. Doktora Tezi, Duke Üniversitesi.
24. İnternet: Plots of pitch mel scale versus Hertz scale. Web: http://en.wikipedia.org/wiki/Mel_scale adresinden 18 Mart 2015'te alınmıştır.
25. Shum, S., Dehak, N., Dehak, R., Glass, J.R. (2010). Unsupervised Speaker Adaptation based on the Cosine Similarity for Text-Independent Speaker Verification. *Speaker and Language Recognition – Scoring, Confidences and Calibration, IEEE Odyssey 2010*.
26. Krause, N., Gazit, R., (2006). SVM-based Speaker Classification in the GMM Models Space. *Speaker and Language Recognition Workshop, IEEE Odyssey 2006*.
27. Lin, T.C., Chen, S.H., Lin, C.C., Truong, T.K. (2006). Speaker Classification Using Support Vector Machine and Wavelets. *IEEE Region 10 Conference, TENCON 2006*, 1-4.
28. Palden Lama, Mounika Namburu, Speech Recognition with Dynamic Time Warping using MATLAB, CS 525, Spring 2010 – Project Report
29. Yang, H.Y., Jing, X.X. (2012). Performance test of Parameters for Speaker Recognition System Based on SVM-VQ, *Machine Learning and Cybernetics (ICMLC), International Conference*, 1,321-325.

30. Fenglei, H., Bingxi, W. (2000). An Integrated System for Text-Independent Speaker Recognition Using Binary Neural Network Classifiers. *Signal Processing Proceedings, WCCC-ICSP 2000, 5th International Conference*, 2, 710-713.
31. Woo, S.C., Lim, C.P., Osman, R. (2001). Devolepment of a Speaker Recognition System Using Wavelets and Artifical Neural Networks. *Intelligent Multimedia, Video and Speech Processing*, 413-416.
32. Cherifa, S., Messaoud, Rç. (2013). New technique to Use The GMM in Speaker Recognition Systems (SRS). *Computer Applications Technology (ICCAT)*, 1-5.
33. Banguoğlu, T. (1986). Türkçenin Grameri, *Atatürk Kültür, Dil Ve Tarih Yüksek Kurumu Türk Dil Kurumu Yayinlari: 528 Türk Tarih Kurumu Basım Evi – Ankara*
34. Özsoy, A.S., Emeksiz, Z.E. (2013). Genel Dilbilim-II, *T.C. Anadolu Üniversitesi Yayini, No: 2429*.
35. Korkmaz, Z. (2003). Türkiye Türkçesi Grameri, *TDK Yayınları, Ankara*.
36. Ergin, M. (1972). Türk Dil Bilgisi, *İstanbul Üniversitesi Edebiyat Fakültesi Yayınları*.
37. Türkçe Sözlük (1955). *Türk Dil Kurumu*, 2. Baskı.
38. Selen, N. (1979). Söyleyiş Sesbilimi, Akustik Sesbilim ve Türkiye Türkçesi. *Türk Dil Kurumu Yayınları, Ankara*.
39. Rosenberg, A.E., Sambur, M.R. (1975). New Techniques for Automatic Speaker Verification. *IEEE Transactions on Aucistics, Speech, and Signal Processing*, 23-2.
40. Linde, Y., Bum, A., ve Gray, R. M. (1980). An Algorithm for Vector Quantizer Design. *IEEE Transactions on Communications*, 28(1), 84-95.
41. İnternet: Wojcicki, K. (2011). Mel frequency cepstral coefficient feature extraction in Matlab. Web:<http://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab/content/mfcc/example.m> adresinden 29 Mayıs 2014'te alınmıştır.
42. İnternet: D. Ellis (2003). Dynamic Time Warp (DTW) in Matlab. Web: <http://www.ee.columbia.edu/~dpwe/resources/matlab/dtw/> adresinden 3 Ağustos 2014'te alınmıştır.

EKLER

EK-1. 12'si erkek 4'ü kadın 16 konuşmacının MFCC öznitelikleri ile sınıflandırıldığında, her kelimenin 2 test verisinin kişilere göre doğru bulunma oranı*

Kişi \	AKTÖR	BARIŞ	TOPRAK	DAMAR	DENİZ	DÜMDÜZ	EKONOMİK	BEŞİK	GİTAR	GURBET	METİN	HÜKÜMET	KAPAK	KARTOPU	KIRMIZI	KÖPÜK
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	0,5	1	1	1	1	0,5	1	1	1	1	0,5	1	0,5	1	1
3	1	1	1	0,5	0	1	1	1	1	0	1	1	0	0,5	0,5	1
4	1	1	1	1	1	1	1	1	1	1	0,5	1	1	1	1	0
5	1	1	1	1	0,5	1	1	1	1	1	1	1	0,5	1	1	1
6	1	1	1	1	1	1	1	0,5	1	1	1	1	1	1	1	1
7	0,5	1	1	1	1	1	1	1	1	1	1	1	0,5	0,5	0	1
8	1	0,5	1	1	1	1	1	1	1	1	1	1	1	1	0,5	1
9	1	1	1	1	1	1	1	1	0	0,5	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	0,5	1	1	0	0,5	1	0,5
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,5
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	0,5	1	0,5	1	0,5	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
% BAŞARI	96,88	93,75	100,00	96,88	87,50	100,00	93,75	96,88	90,63	87,50	96,88	96,88	81,25	87,50	87,50	87,50

- * 1 : 2 test verisinin de o kişi ve o kelime için sistem tarafından doğru bilinmesi durumu
0,5 : 2 test verisinden o kişi ve o kelime için sistem tarafından birinin doğru bilinmesi durumu
0 : 2 test verisinden o kişi ve o kelime için sistem tarafından hiçbirinin doğru bilinmemesi durumu

EK-1. (devam) 12'si erkek 4'ü kadın 16 konuşmacının MFCC öznitelikleri ile sınıflandırıldığında, her kelimenin 2 test verisinin kişilere göre doğru bulunma oranı*

Kişi \	TIRNAK	TÜTÜN	TUZAK	BOZUK	BÖREK	PAZAR	PERŞEMBE	BURUŞUK	TAM	TEKRAR	BÜTÜN	BIÇAK	BİDON	DUMAN	DİLEK	KADIN
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	0,5	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	0,5	1	1	0,5	0,5	1	1	1	1	1	1
4	1	1	1	1	1	1	0,5	1	1	1	0,5	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	0,5	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	0,5	0,5	1	1	1	1	1	1	1	0,5	1	0,5
10	1	1	1	1	1	1	1	1	1	0,5	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	0,5	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	1	0,5	1	1	1	1	1	1	0,5	1	1	1	0,5	1	1	1
16	1	0,5	1	1	1	1	1	1	1	1	1	1	1	1	1	1
% BAŞARI	100,00	93,75	100,00	100,00	96,88	87,50	96,88	96,88	93,75	93,75	96,88	100,00	96,88	96,88	100,00	96,88

- * 1 : 2 test verisinin de o kişi ve o kelime için sistem tarafından doğru bilinmesi durumu
0,5 : 2 test verisinden o kişi ve o kelime için sistem tarafından birinin doğru bilinmesi durumu
0 : 2 test verisinden o kişi ve o kelime için sistem tarafından hiçbirinin doğru bilinmemesi durumu

EK-1. (devam) 12'si erkek 4'ü kadın 16 konuşmacının MFCC öznitelikleri ile sınıflandırıldığında, her kelimenin 2 test verisinin kişilere göre doğru bulunma oranı*

kişi \	DOLAP	DÖŞEK	KEDİ	KUZU	KIRAZ	GARANTİ	GENİŞ	GOFRET	GÖMLEK	GIDA	GÜNEŞ	POLİS	PÖRSÜK	PIRASA	PİLAV	PÜRE
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	0,5	1	1	1	1	1	1	1
4	1	1	0,5	1	1	1	1	0,5	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	0,5	1	0,5	1	1	1	1	1	1	1	1	1	1	1	1
10	0,5	0,5	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	0,5	0,5	0,5	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,5	1
16	1	1	1	1	0,5	1	1	1	1	1	1	1	1	1	0	1
% BAŞARI	96,88	93,75	96,88	96,88	93,75	96,88	96,88	96,88	96,88	100,00	100,00	100,00	100,00	100,00	90,63	100,00

- * 1 : 2 test verisinin de o kişi ve o kelime için sistem tarafından doğru bilinmesi durumu
0,5 : 2 test verisinden o kişi ve o kelime için sistem tarafından birinin doğru bilinmesi durumu
0 : 2 test verisinden o kişi ve o kelime için sistem tarafından hiçbirinin doğru bilinmemesi durumu

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : ŞENTÜRK, Zekeriya
 Uyruğu : T.C.
 Doğum tarihi ve yeri : 27.05.1987, Arhavi
 Medeni hali : Evli
 Telefon : 0 (543) 256 03 08
 Faks : -
 e-mail : zsenturk@kho.edu.tr



Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Yüksek lisans	Gazi Üniversitesi / Elektrik - Elektronik	2015
Lisans	Karadeniz Tek. Üni./ Elektrik - Elektronik	2011
Lise	Arhavi Süper Lisesi	2005

İş Deneyimi

Yıl	Yer	Görev
2011-Halen	Kara Harp Okulu	Araştırma Görevlisi

Yabancı Dil

İngilizce

Yayınlar

- ‘Özilinti Matrisi Kullanılarak Ötümlü-Ötümsüz Konuşma Sınıflandırma’, SİU 2014
- ‘LSB Yöntemiyle Renk Değişimi Esaslı Yeni Bir Steganografik Yaklaşım’, SİU 2015

Hobiler

Yüzme, Koşu, Bilardo, Balık Tutmak