

Ortak Vektör Yaklaşımı İle Fonem Tabanlı
Türkçe Yalıtık Kelime Tanıma

Serkan Keser

YÜKSEK LİSANS TEZİ

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Ağustos 2008

Phoneme Based Turkish Isolated Word Recognition
Using Common Vector Approach

by
Serkan Keser

MASTER OF SCIENCE THESIS

Department of In Electrical-Electronic Engineering

August 2008

Ortak Vektör Yaklaşımı İle Fonem Tabanlı
Türkçe Yalıtık Kelime Tanıma

Serkan Keser

Eskişehir Osmangazi Üniversitesi
Fen Bilimleri Enstitüsü
Lisansüstü Yönetmeliği Uyarınca
Elektrik ve Elektronik Mühendisliği Anabilim Dalı
Telekomünikasyon Bilim Dalında
YÜKSEK LİSANS TEZİ
Olarak Hazırlanmıştır

Danışman: Yrd.Doç. Dr. Rifat EDİZKAN

Ağustos 2008

Elektrik-Elektronik Mühendisliđi Anabilim Dalı Yüksek Lisans öđrencisi Serkan Keser'in YÜKSEK LİSANS tezi olarak hazırladıđı "Ortak Vektör Yaklaşımı İle Fonem Tabanlı Türkçe Yalıtık Kelime Tanıma" başlıklı bu çalışma, jürimizce lisansüstü yönetmeliđin ilgili maddeleri uyarınca deđerlendirilerek kabul edilmiştir.

Üye : Yrd.Doç.Dr. Rifat EDİZKAN (Danışman)

Üye : Prof.Dr. M.Bilginer GÜLMEZOĐLU

Üye : Yrd.Doç.Dr. Atıf ÇAY

Üye : Prof.Dr. Atalay BARKANA

Üye : Doç.Dr. Ömer Nezih GEREK

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun tarih ve sayılı kararıyla onaylanmıştır.

Prof. Dr. Nimetullah BURNAK

Enstitü Müdürü

ÖZET

Bu tez çalışmasında, Ortak Vektör Yaklaşımı (OVY) ile fonem tabanlı Türkçe yalıtık kelime tanıma üzerinde çalışılmıştır. OVY, kelime tabanlı ses tanımda yüksek tanıma başarımları verdiği bilinmektedir. Fakat bir dildeki ses birimlerinin süresinin kısa olması, bu birimlerin bazılarının çok sık kullanılmaması ve özellik uzayında bunların dağılımlarının iç içe olması ses birim tabanlı ses tanıma başarımlarını etkilemektedir. Tez çalışmasında sınıf modelleri ve tanıma başarımları METU veri tabanı üzerinden elde edilmiştir. Tanınacak kelime üzerinde 10 ms aralıklarla pencere kaydırılarak her bir aralık için ses birim sınıflaması yapılmış ve daha sonra elde edilen harf dizisi Redundant Hash Addressing (RHA) uygulanarak dizinin en yakın olduğu kelime bulunmuştur. Sınıflama hatalarından kaynaklanan yanlış tanıma sayısını azaltmak için RHA'de iyileştirmeler yapılmıştır. Tez çalışmasında kişi bağımlı ve kişi bağımsız kelime tanıma başarımları elde edilmiştir. OVY ile fonem tanımda eğitim kümesinde %99 tanıma oranı elde edilmiştir. Test kümesinde kişi bağımsız fonem tanıma başarımları ünlü ve ünsüz sesler için sırasıyla %48 ve %58 civarındadır. RHA'de yapılan iyileştirmeler sonucunda kişi bağımsız ve bağımlı kelime tanıma başarımları sırasıyla %72.22 ve %82.5 olarak elde edilmiştir. Fonemlerin sınıflama başarımları düşük olmasına rağmen, RHA kullanılarak yüksek kelime tanıma başarımları elde edilmiştir. Yapılan çalışmaların sonuçları sınıflama başarımlarını arttırmak için OVY'de ve kelime çözümlemesinde iyileştirmeler yapılması gerektiğini göstermektedir.

Anahtar Kelimeler: Ortak Vektör Yaklaşımı, fonem tanıma, yalıtık ses tanıma,
Redundant Hash Adresleme

SUMMARY

In this thesis, phoneme based Turkish isolated word recognition is studied using Common Vector Approach(CVA). It is known that CVA provides successful results in word based speech recognition. On the other hand, the success of phoneme-based speech recognition is influenced due to the fact that the phonemes in a language are short; some of them are not used frequently and their distributions in the feature space overlap. In the thesis study, the class models and recognition rates have been obtained from METU database. Phoneme classification has been performed on the speech windows with 20 ms duration. The speech window is moved by 10 ms to classify the succeeding phonemes. Once the phoneme series is obtained, then the closest word in the dictionary has been found by applying Redundant Hash Addressing (RHA). In order to decrease the amount of inaccurate recognition due to classification errors, certain improvements have been applied in RHA. In the thesis study, speaker-dependent and speaker- independent word recognition has been achieved. The success in phoneme recognition using CVA in the training set is over 95%. The correct recognition rates in speaker- independent phoneme recognition in the training set are approximately 48% for vowels and 58% for consonants. The speaker-independent and speaker dependent recognition rates are increased to 72.22% and 82.5% respectively as a result of improvements in RHA. Despite the success of phoneme classification is low, high rates of word recognition have been achieved by using the RHA. The results of this study point out that it is necessary to improve CVA and the word decoder in order to increase the recognition rates.

Keywords: Common Vector Approach, phoneme recognition, isolated word recognition, Redundant Hash Addressing.

TEŞEKKÜR

Tüm Yüksek Lisans Programı boyunca desteğini hiçbir zaman esirgemeyen, danışman hocam Yrd.Doç.Dr.Rifat Edizkan'a, ve tüm Eskişehir Osmangazi Üniversitesi Elektrik-Elektronik Mühendisliği Bölümüne teşekkürlerimi sunarım.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	v
SUMMARY	vi
TEŞEKKÜR	vii
ŞEKİLLER DİZİNİ	x
TABLolar DİZİNİ	xi
KISALTMALAR DİZİNİ	xii
EKLER	xiii
1. GİRİŞ VE AMAÇ	1
2. FONEM TABANLI YALITIK KELİME TANIMA	4
2.1 Fonem Tabanlı Ses Tanımaya Genel Bakış.....	4
2.2 Türkçe Sesbirim Yapısı.....	5
2.2.1 Parça sesbirimlerle ilgili ayırıcı özellikler.....	6
2.2.2 Parçalar üstü sesbirimlerle ilgili ayırıcı özellikler.....	8
2.2.3 Türkçe fonem yapısı.....	9
3. ORTAK VEKTÖR YAKLAŞIMI	11
3.1 Veri Tabanı.....	11
3.2 Fonem Uzunluklarının Eşitlenmesi.....	12
3.3 Parametrelerin Elde Edilmesi.....	13
3.3.1 Çerçeveleme.....	13
3.3.2 Pencereleme.....	13
3.4 Ortak Vektör Yaklaşımı.....	15

İÇİNDEKİLER (Devam)

Sayfa

4. REDUNDANT HASH ADRESLEME	18
4.1 Hash Fonksiyonu.....	19
4.2 Hash Tablosu	19
4.3 Çatışmanın Çözümlemesi	20
4.3.1 Açık adresleme	21
4.3.2 Bağlı liste kullanımı.....	21
4.4 Hash Sözlükten Uygun Kelimenin Seçimi.....	21
5. DENEYSEL ÇALIŞMALAR.....	24
5.1 Parametre Seçimi.....	25
5.2 Fonem Sınıflama Başarımları.....	27
5.3 Kelime Çıkartma	31
5.4 RHA ile Kelime Çözümleme.....	39
6. SONUÇ VE ÖNERİLER	43
7. KAYNAKLAR DİZİNİ	59

ŞEKİLLER DİZİNİ

<u>Sekil</u>	<u>Sayfa</u>
3.1 Fonem sınıfları için ortak vektörün hesaplama aşamaları.....	11
3.2 Fonem öznitelik vektör uzunluklarını eşitleme yöntemleri	12
3.3 Kısa fonemler için öznitelik vektörlerinin boyunun eşitlenmesi	13
3.4 Mel bank filtre yapısı.....	14
4.1 RHA metodunun gösterimi.....	18
4.2 Hash tablosu.....	19
4.3 Açık adresleme örneği	21
4.4 Bağlı liste ile çatışmanın aşılması	21
4.5 Hash adresleme	22
5.1 METU veritabanı ortalama fonem uzunlukları.....	25
5.2 Çalışmada kullanılan pencere yapısı	26
5.3 10 ünlü fonem sınıfı için farklı sürelerde bulunan başarımlar.....	29
5.4 22 ünsüz fonem sınıfı için farklı sürelerde bulunan başarımlar.....	29
5.5 8 ünlü fonem sınıfı için farklı sürelerde bulunan başarımlar.....	30
5.6 19 ünsüz fonem sınıfı için farklı sürelerde bulunan başarımlar.....	31
5.7 ‘Gülücük’ kelime verisi için enerji ve sıfır geçiş oranı.....	31
5.8 ‘Gülücük’ kelime verisi için düzgünleştirilmiş enerji ve sıfır geçiş oranı.....	32
5.9 ‘Saticılık’ kelimesi için eşik değeri ile fonem sınıflama.....	32
5.10 ‘sıyaf’ karakter dizisinin iyileştirilmesi.....	35
5.11 ‘Hayat’ kelimesi için ünsüz fonem sınırlarının geniş alınması.....	36
5.12 ‘Şehir’ kelimesi için bağlı liste ile kelime çözümleme.....	39
5.13 RHA ile yapılan iyileştirme algoritması	41

TABLÖLAR DİZİNİ

<u>Cizelge</u>	<u>Sayfa</u>
2.1 IPA, SAMPA ve METUbet fonem sembolleri.....	9
2.2 METUbet fonem sembolleri.....	10
5.1 METU veri tabanı fonem sayıları.....	24
5.2 80 ms pencere uzunluđu ile kelime çıkartma	37
5.3 60 ms pencere uzunluđu ile kelime çıkartma.....	38
5.4 100 ms pencere uzunluđu ile kelime çıkartma.....	38
5.5 120 ms pencere uzunluđu ile kelime çıkartma.....	38
5.6 150 ms pencere uzunluđu ile kelime çıkartma.....	39

KISALTMALAR

<u>Kısaltmalar</u>	<u>Açıklama</u>
CVA	Common Vector Approach
DTW	Dynamic Time Warping
FFT	Fast Fourier Transform
HMM	Hidden Markov Model
IPA	International Phonetic Alphabet
IVR	Interactive Voice Responce
LPC	Linear Predictive Coefficient
METU	Middle East Technical University
METUbet	Middle East Technical University Alphabet
MFCC	Mel Cepstrum Cepstral Coefficient
NN	Neural Network
PLP	Perceptual Linear Predictive Coefficients
RHA	Redundant Hash Adresleme
SAMPA	Speech Assessment Method Phonetic Alphabet
OVY	Ortak Vektör Yaklaşımı

EKLER

	<u>Sayfa</u>
EK-A Yöntem 2A ile kişi bağımlı kelime tanıma.....	45
EK-B Yöntem 2A ile kişi bağımsız kelime tanıma.....	47
EK-C Yöntem 2A ve 2B ile kişi bağımsız kelime tanıma.....	51
EK-D Yöntem 2A ve 2B ile kişi bağımlı kelime tanıma.....	55
EK-E 60, 80, 100,120, 150 ms uzunluklu 32 fonem sınıfı için bulunan başarımlar....	57
EK-F 60, 80, 100,120, 150 ms uzunluklu 27 fonem sınıfı için bulunan başarımlar ...	58

BÖLÜM 1

GİRİŞ

Ses tanıma uygulamaları günümüzde değişik alanlarında kullanılmaktadır. Bu alanlara; sesli komut uygulamalarında (robot, taşıt ve bilgisayar kontrolü), haberleşme teknolojilerinde kullanılan IVR (Interactive Voice Responce) sistemlerinde, sesi yazıya dönüştürme, konuşmacı tanıma ve onaylama, anahtar kelime tanıma örnek olarak verilebilir (Edizkan R. vd., 2007; Bayrakçeken, M.K.,2005; Furui, S et. al., 2004).

Ses tanımada Dynamic Time Warping (DTW), Hidden Markov Model (HMM), Neural Networks (NN) veya altyuzay sınıflandırıcılar (FLDA, CLAFIC, CVA) kullanılabilir. Bu yöntemler içinde “state of art” olarak bilenen HMM, ses tanıma sistemlerinde yaygın olarak kullanılmaktadır. Bunun yanında HMM ile karşılaştırılabilir başarımlar veren yöntemlerin ses tanımada kullanımı araştırmacıların ilgisini çekmektedir (Gülmezoğlu,M.B., vd.,1999 ; Gorin, A. and Mammone, R.J., 1994 ; Abdulla, W.H., et al.,2003). Bunlar içinde bir altuzay sınıflandırma yöntemi olan Ortak Vektör Yaklaşımı (OVY), sınırlı sayıda yalıtık kelime tanıma uygulamalarında HMM’le karşılaştırılabilir başarımlar vermektedir (Gülmezoğlu,M.B., vd.,1999, 2001, 2007; Gunal, S., 2007; Gülmezoğlu,M.B.,, et al., 2007). Ayrıca OVY sınıflandırma yöntemindeki basitlikten dolayı HMM göre daha hızlı çalışmaktadır. OVY yöntemi ses tanıma yanında, konuşmacı tanıma ve anahtar kelime tanıma uygulamalarında kullanılmıştır (Bayrakçeken, M.K., 2005).

OVY’nin ses tanıma uygulamalarında şimdiye kadar kelime tabanlı sınıflandırma yapılmıştır. Yapılan çalışmalarda sınırlı sayıda kelime kullanılarak %95’ in üzerinde başarımlar elde edilmiştir. OVY’nin çok sayıda kelime tanıma için başarımının ne olacağı bu konuda çalışma yapılmadığı için bilinmemektedir.

Kelime tabanlı ses tanımada her kelime için veri tabanı ihtiyacı vardır. Bu nedenle kelime tabanlı tanıma sistemlerine yeni kelime eklemek her zaman kolay olmayabilir. Çünkü istenilen kelime için veri tabanı hemen elde edilemeyebilir. Bu da

ses tanıma sistemin ticari olarak kullanılmasını güçleştirir. Fakat fonem tabanlı ses tanıma yapılan sistemlerde fonem dizileri kullanılarak tanıyıcı sisteme yeni bir kelime eklemek mümkündür. Bu da tanıyıcı sistemin kullanılabilirliği arttırmaktadır.

Fonem tabanlı ses tanıma sistemlere bir dildeki her fonem sınıfı için bir model oluşturulur. Bu sistemlerde yüksek başarımlar elde etmek için kullanılan veri tabanının dildeki bütün farklı söyleyişleri kapsamı gerekir. Veri tabanında fonem sınırlarını kesin olarak belirlemek çok zordur. Bunun yanında ses yolunun farklı şekil almasında dolayı fonemler farklı şekilde üretilebilmektedir. Bütün bunlar fonem dağılımlarının iç içe geçmesine neden olmakta ve sınıflandırma başarımlarını düşürmektedir. Ayrıca kelimelerin fonetik gösterimlerindeki bazı fonemler coarticulation'dan dolayı konuşmacı tarafından yutulabilmektedir. Bu da fonem dizilerinden kelime tanımayı etkileyen başka bir etkidir. Fonemler kısa süreli olmalarından dolayı çok fazla bilgi içermemektedirler. Bu nedenlerden dolayı fonem sınıflandırma başarımları çok yüksek değildir (Arısoy, E. ve Arslan, L.M 2005). Fakat fonem tanıma sonuçları dil modeliyle birleştirilerek çok yüksek kelime tanıma oranları elde edilebilmektedir (Arslan, L.M, 1999; Karjanadecha M. and Zahorian, S., 2001; Turk, O. ve Arslan, L.M., 2004; Kurcan et al., 2006). Literatürde daha çok üçlü ses birimi (triphone) kullanılarak ses tanıma yapılmaktadır. Üçlü ses birimleriyle dildeki farklı söyleyişler modellenebilir. Bir dilde sayıları binlerle ifade edilen üçlü ses birimleri bulunmaktadır. Yapılan çalışmalara göre Türkçe'de en sık kullanılan üçlü ses birim sayısı 1000 civarındadır.

Bu çalışmada OVY ile fonem tabanlı Türkçe yalıtık kelime tanıma konusu üzerinde çalışılmıştır. Burada, OVY'nin fonem tabanlı sınırlı sayıda kelime tanımda HMM' ye alternatif bir yöntem olarak kullanılabileceğini göstermek amaçlanmıştır. Türkçe dili METUbet fonetik alfabesine göre 39 fonem sınıfından oluşmaktadır (Salor, Ö. et al., 2002). Fonem sınıflarının modellerinin oluşturulmasında ve kelime tanımda METU veri tabanı kullanılmıştır. Bu çalışmada fonem tabanlı kelime tanıma başarımları kişi bağımlı ve kişi bağımsız olarak elde edilmiştir. Fonem sınıflama, kelime verisi üzerinde 80 ms (1280 örnek)'lik pencerenin her 10 ms aralıklarla kaydırılması ile yapılmıştır. Bulunan fonem dizisi kullanılarak ilgili kelimeye ait karakter dizisi

bulunmuştur. Bu karakter dizisi kullanılarak Redundant Hash Addressing (RHA) ile kelime çözümlemesi yapılmıştır .

Bu çalışmada OVY bilinen klasik şekliyle fonem tanımada kullanılmıştır. Fonem dizilerinin RHA ile çözümlenmesindeki başarımı artırmak için çeşitli yöntemler kullanılmıştır.

Tezin çalışmasının düzeni şu şekildedir:

Birinci bölümde fonem tabanlı yalıtık kelime tanıma başlığı altında ses tanıma sistemlerinin sınıflandırılması ile Türkçe fonetik yapısı anlatılmaktadır. İkinci bölümde OVY'nin yeterli ve yetersiz veri durumunda nasıl uygulandığı verilmektedir. Üçüncü bölümde ise fonem dizilerinde kelime çözümleme tekniği olarak kullanılan RHA anlatılmıştır. Deneysel çalışma sonuçları ve tez çalışmasından elde edilen sonuç ve öneriler sırasıyla dördüncü ve beşinci bölümde verilmektedir.

BÖLÜM 2

FONEM TABANLI YALITIK KELİME TANIMA

Sesbirimler (fonlar) sesin en küçük birimleridir. Sesbirimlere karşılık gelen sembollere fonem adı verilmektedir. Sesbirimler ses organlarının farklı kullanım biçimlerine göre değişik özellikte oluşabilmektedir. Bu sesbirimlere (fonlara) allofon denmektedir (Salor, Ö. et al., 2002).

Seslerin tanınması genellikle sözcük ya da fonem tabanlı olarak yapılmaktadır. Sözcük tanıma için o dili oluşturan çok sayıda kelime olduğundan işlem yükü fonem tabanlı tanımaya göre oldukça fazladır. Fonem tabanlı kelime tanıma için ise o dili oluşturan az sayıda fonem olduğundan işlem açısından daha hızlı olduğu açıktır. Ayrıca fonem tabanlı tanıma sisteme istenildiği kadar farklı kelime eklenerek, bu kelimeler için eğitim yapılmaksızın tanıma işlemi yapılabilir.

2.1 Fonem Tabanlı Ses Tanımaya Genel Bakış

Günümüzde ses tanıma uygulamaları pek çok alanda kullanılmaktadır. Sesle cihaz kontrolü, konuşmacı tanıma ile kimlik tespiti, bilgisayar ortamında sesi yazıya çevirme vb. uygulamalar sayılabilir. Ses tanıma sistemleri konuşmacıya bağımlılık, sesin sürekliliğine ve tanımda kullanılacak seslerin uzunluklarına göre sınıflara ayrılır (Furui, 1989).

Ses tanıma konuşmacıya bağımlı veya bağımsız olarak yapılabilir. Konuşmacı bağımlı ses tanıma uygulamalarında, bir kişi tarafından seslendirilmiş ses verileri kullanılarak o kişi için tanıma yapılır. Konuşmacı bağımsız ses tanımda ise farklı kişilerce seslendirilmiş ses verilerinden faydalanılarak herhangi bir ses tanıma çalışılır.

Kelimeler arasında durarak konuşma veya doğal konuşmanın tanınmasında sırasıyla yalıtık ve sürekli ses tanıma yöntemleri kullanılır. Yalıtık ses tanıma yöntemleri küçük sözlüğe sahip tanıma uygulamalarında tercih edilmektedir. Yalıtık kelime tanımada, seslerin başlangıç ve bitiş yerleri kolay bulunabildiği için elde edilen başarımlar yüksektir. Bu yöntemin dezavantajı, sözlükte değişiklik yapmaya her zaman imkan vermemesidir. Bu nedenle, tanıma sistemini değişik uygulamalarda kullanmak ancak ses veri tabanı varsa mümkün olmaktadır.

Sürekli ses tanıma yöntemlerinde sesbirimleri kullanılarak kelime ve kelimelerden cümle çözümlemesi yapılmaktadır. Bir dildeki sesbirimlerinin sürelerinin kısa olması ve ses içindeki yerlerinin tam olarak belirlenememesinden dolayı, sesbirimlerinin sınıflama başarımı çok yüksek değildir (Kohonen, T., et al., 1980). Fakat sesbirimleriyle birlikte dil modellemesi kullanılarak yüksek kelime tanıma başarımları elde edilebilmektedir. Bunun için dildeki sesbirimlerinin kullanılma ve bunların birbirlerini takip edebilme istatistiklerinin çıkartılması gerekir (Arisoy, E., et al., 2006). Sesbirimleriyle tanıma yapılan sistemlerde kullanılan sözlük üzerinde değişiklik yapmak kolay olmaktadır. Bu sistemleri değişik uygulama alanlarında kullanmak mümkün olabilmektedir.

2.2 Türkçe Sesbirim Yapısı

Sesler anlamlı birimleri oluştururlar. Anlamlı birimler içerisinde seslerin iki türlü özelliği vardır: ‘ayırıcı özellik’ , ‘çevresel özellik’ (Selen,N., 1979). Ayırıcı özellikler, eş uzunlukta iki birimin anlamını birbirinden ayıran özelliklerdir. Örneğin ‘sin’ ve ‘sün’ sözcükleri ünlülerin yuvarlak olup olmamasıyla; ‘sor’ ve ‘zor’ sözcükleri ise ses tellerinin başta titreşip titreşmemeleriyle ayrılırlar. ‘sin’ ve ‘sün’ sözcüklerindeki /s/ sesi ‘sin’ için düz ‘sün’ için yuvarlak olarak farklı özellikte olmalarına rağmen anlam ayırımında katkıları bulunmaz. /s/ seslerinin düz ya da yuvarlak oluşları buldukları yere bağlı çevresel özelliklerdir.

Ayrıcı özellikleri özdeş, çevresel özellikleri bakımından ayrı olabilen sesler topluluğuna fon ya da sesbirim adı verilmektedir. Türkçe’de sesbirimler (fonlar) ikiye ayrılmaktadır. Bunlar ;

- Parça sesbirimler
- Parçalar üstü sesbirimlerdir.

Parçalı sesbirimler fonemlerle ifade edilebilen sesbirimlerdir. Her bir parçalı sesbirime ait bir fonem vardır. Bu yüzden sesli ifade tanıma uygulamalarında parça sesbirimler kullanılmaktadır. Parçalar üstü sesbirimler ise sesin vurgu, uyum, perde değişimi ve süresi ile ilgilidir.

2.2.1 Parça sesbirimlerle ilgili ayırıcı özellikler

- **Ses tellerinin titreşimi**

Türkçede ses tellerinin titreşimine göre sesbirimler sınıflandırılabilir.

Ünlüler : a, e, ı, i, o, ö, u, ü

Ünsüzler: b, c, ç, d, f, g, ğ, h, j, k, l, m, n, p, r, s, ş, t, v, y, z

Türkçedeki ünlüler ses tellerinin titreşimi ile oluşurlar. Ünsüzlerin ise bir kısmında titreşim varken (ötümlüler) bir kısmında yoktur(ötümsüzler) .

Ötümlüler : b, d, g, v, z, j, c, ğ, l, r, m, n, y

Ötümsüzler: p, t, k, f, s, ş, ç, h

- **Sesi çıkaran organlar ve seslerin çıkış yerleri**

Seslerin çıkış yerleri ve ses organlarının konumuna göre sınıflandırma yapılırken ünlü ve ünsüzler ayrı olarak incelenmektedir. Ünlülerde çıkış yeri dilin ağızda en yüksek yerine göre değişmektedir.

Dil önü ünlüleri : i, ö, e, ü

Dil arkası ünlüleri : ı, u, a, o

Ünsüzler için ise çıkış yerlerinin yanı sıra yardımcı ses organlarının etkileşimi de belirleyici bir özelliktir.

Çift dudak	: p, b, m
Alt dudak /Üst dişler	: f, v
Dil ucu /Diş arkası	: t, d
Dil ucu /Diş eti	: s, z, n, r, l, ç, c
Dil Önü /Sert damak	: s, j, y
Dil arkası/damak	: k, g
Ses teli /Gırtlak	: h

- **Çıkış biçimleri**

Seslerin çıkış biçimlerine göre sınıflandırmada ses organlarının ses çıkarırken aldıkları durumlar belirleyicidir. Bu sınıflandırmada da ünlü ve ünsüzler için ayrı inceleme yapılmaktadır. Ünlüler için dil/ damak arası açıklığına göre;

Dar : i, ü, ı, u

Orta : e, ö, - , o

Geniş : -, - , a, -

Dudakların biçimine göre ;

Düz : ı, i, e, a

Yuvarlak : ü, u, ö, o

sınıflanmaktadır. Ünsüzler öncelikle ağız ve geniz ünsüzleri olmak üzere ikiye ayrılmaktadır.

Ağız : b, c, ç, d, f, g, h, j, k, l, p, r, s, s, t, v, y, z

Geniz : m, n

Ağız ünsüzleri de patlamalı, sızmalı, patlamalı-sızmalı, yan, çarpmalı ve yarı ünlü olarak sınıflandırılabilir.

Patlamalı ünsüzler;

Çift dudak : b, p

Dil ucu/ Diş arkası : d, t

Dil / Damak : g, k

Sızmalı ünsüzler

Dudak / Dis : v, f

Dil ucu/ Dis eti : z, s

Dil önü/ Serdamak : j, s

Sestelleri : h

Patlamalı sızmalı ünsüzler

Dil ucu/ Dis eti : c, ç

Yan ünsüz : l

Çarpmalı ünsüz : r

Yarı ünlü : y

2.2.2 Parçalar üstü sesbirimlerle ilgili ayırıcı özellikler

Parçalar üstü sesbirimler, süre, soluk baskısı, uyum, perde değişimi ve ezgi açılardan birbirinden ayrılırlar.

- Süre: Ünlülerin uzun ya da kısa olması anlamında kullanılmaktadır.

Örneğin : ‘eğ’ kelimesinde ‘e’ uzun, ‘dek’ kelimesinde ise kısadır.

- Soluk baskısı (vurgu) : Konuşurken, fizyolojik olarak hece çıkararak göğüs atışlarını güçlendirici atışlar yapılır. Bu atışlar vurgulu hecelere denk gelir.

Örneğin : ‘konuşma’ kelimesi için son hece vurgulanırsa (ma) ‘sus konuşma artık’ cümlesindeki anlam, ikinci hece vurgulanırsa (‘konuşma’), ‘Ali bey bir konuşma yapacak’ cümlesindeki başka bir anlam oluşacaktır.

- Uyum;

Dil açıklığı: dar-genis

Çıkış noktası: ön-arka

Dudak biçimi: düz-yuvarlak

Burada belirtilen karşıtlıkların sözcük içinde gerçekleşmeyeceğine dayanan bir sınıflandırmadır. Örneğin sözcük içinde düz ünlülerden sonra yuvarlak ünlü kullanılmaz.

- Perde Değişimi: Sözcük içinde bir tek hecenin ses düzeyinin değişimine göre de anlam farklılığı oluşmaktadır. Örneğin : ‘ya, (gördün mü?)’ ile ‘ya, (öyle mi?)’ cümlelerindeki ‘ya’ hecesinin farklı ses düzeyleri kullanıldığında farklı anlamlar oluşmaktadır.
- Ezgi : Bir tümceyi örten perde değişiminin tümü o tümcenin ezgisidir.

2.2.3 Türkçe fonem yapısı

Modern standart Türkçe, fonem tabanlı bir dildir. Yani her fonem, yazı dilindeki bir harfle temsil edilebilmektedir. Orta Doğu Teknik Üniversitesinde ki bir grup tarafından yapılan çalışmada Türkçe yazı dilindeki 29 harfe karşılık, konuşma dilinde (allofonlar ile) 45 sesbirim ve bu sesbirimlere karşılık gelen 39 fonem sınıfı olduğu tespit edilmiştir (Salor, Ö. et al., 2002). Otuzdokuz (39) fonem sınıfını içeren METUbet (Middle East Technical University Alphabet) fonetik alfabe oluşturularak, ses tanıma uygulamalarında kullanılması sağlanmıştır.

IPA	SAMPA	METUbet
ɑ	A	AA
a	a	A
e	e	E
ɛ	E	EE
i	i	IY
ɪ	I	IY
ı	ı	I
ɔ	O	O
o	o	O
U	U	U
u	u	U
œ	2	OE
ø	5	OE
Y	Y	UE
y	y	UE
b	b	B
d	d	D
g	G	GG
ʃ	g	G
h	h	H
ʒ	Z	J
k	k	KK
c	c	K
l	L	L
l̥	l	LL
m	m	M
n	n	NN
ŋ	N	N
p	p	P
r	r	R
ʀ	R	RR
ɻ	4	RH
s	s	S
ʃ	S	SH
t	t	T
v	v	VV
ʋ	w	V
j	j	Y
ɹ	yy	Y
z	z	Z
ʒ	zz	ZH
ʒ	DZ	C
ʃ	TS	CH
f	f	F
:	:	GH

Tablo 2.1 IPA, SAMPA ve METUbet fonem sembolleri

Tablo 2.2’de görüldüğü gibi Türkçe ses tanıma uygulamalarında IPA (International Phonetic Alphabet) ya da SAMPA (Speech Assessment Method Phonetic Alphabet) fonem sembollerinin kullanımı zor olduğundan SAMPA alfabesindeki 45 fonem için, METUbet fonetik alfabede 39 fonem sembolü kullanılmıştır. Bu fonem sınıfları aşağıdaki fonem sembollerinden oluşmaktadır.

A	AA	B	C	CH	D	E	EE	F	G	GG	GH	H	I	IY	J	K	KK	L	LL
M	N	NN	O	OE	P	R	RH	RR	S	SH	T	U	UE	V	VV	Y	Z	ZH	

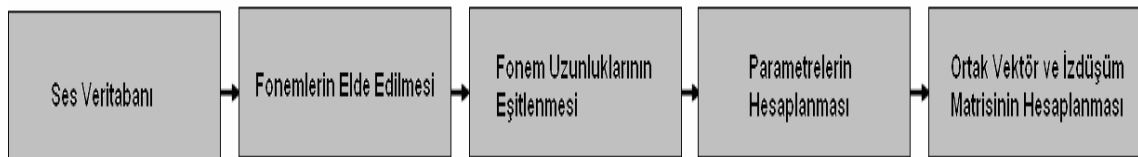
Tablo 2.2 METUbet fonem sembolleri

BÖLÜM 3

ORTAK VEKTÖR YAKLAŞIMI

OVY her sınıfa ait değişmeyen özelliklerin, ortak bir vektörle ifade edildiği bir altuzay sınıflama şeklidir. Söylenen seslerdeki şive, cinsiyet, hastalıklardan kaynaklı ses bozuklukları gibi durumlardan oluşan farklılıklar atılarak her sesin ortak özelliklerini içeren tek bir vektör elde edilir (Gülmezoğlu, M.B. vd.,1999, 2001). Ortak vektör yaklaşımı verinin yetersiz ve yeterli olduğu durumları için kullanılabilir. Yetersiz veri durumunda her sınıf için kullanılan öznitelik vektörleri sayısı (m), öznitelik vektörlerinin boyutlarından (n) küçük olmaktadır. Yeterli veri durumu için ise kullanılan öznitelik vektörleri sayısı (m), öznitelik vektörlerinin boyutlarından (n) büyük ya da eşit olmaktadır (Gülmezoğlu, M.B et al, 2007).

OVY eğitim aşamasında, fonem sınıflarına ait verilerin uzunlukları eşitlenir, öznitelik parametreleri çıkartılır ve en son olarak ortak vektör ve farksızlık uzayı izdüşüm matrisi elde edilir. Şekil 3-1 'de Ovy 'deki işlem basamakları gösterilmektedir.



Şekil 3.1 Fonem sınıfları için ortak vektörün hesaplama aşamaları

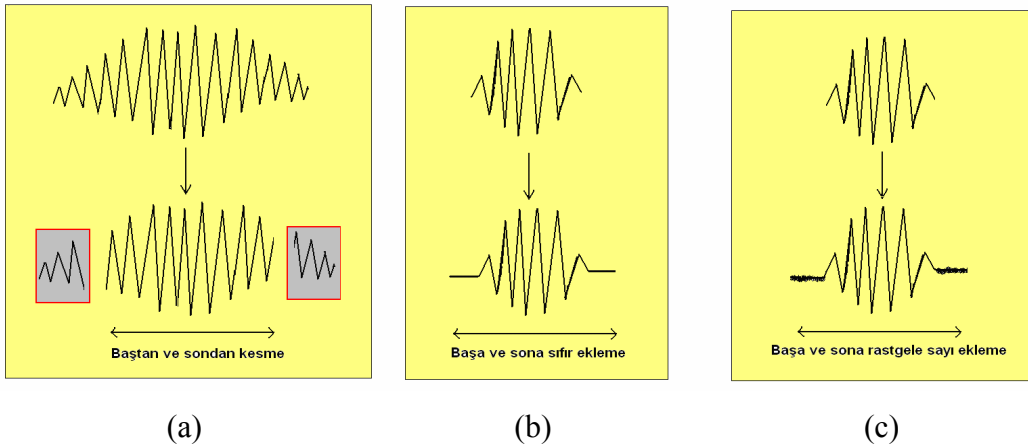
3.1 Veri tabanı

Fonem tabanlı Türkçe yalıtık kelime tanıma çalışması için Orta Doğu Teknik Üniversitesi'nde oluşturulan, Türkçe ses özelliklerini en geniş biçimde kapsayacak şekilde tasarlanmış METU veritabanı kullanıldı. Bu veritabanında Türkçe ses verilerine ait kelime ve fonem başlangıç-bitiş değerleri etiketlenmiş halde dosyalarda

bulunmaktadır. Veritabanı, 120 konuşmacıdan alınan 40'ar adet üçlü ses dengeli cümlelerin seslendirilmesi ile Türkçe'de 29 harfe karşılık gelen 38 sesbirim göz önünde bulundurularak oluşturulmuştur. Burada 'ğ' harfine karşılık gelen sesbirim kullanılmadığından, METUbet fonetik alfabe 38 fonem sınıfı ile temsil edilmişlerdir.

3.2 Fonem Uzunluklarının Eşitlenmesi

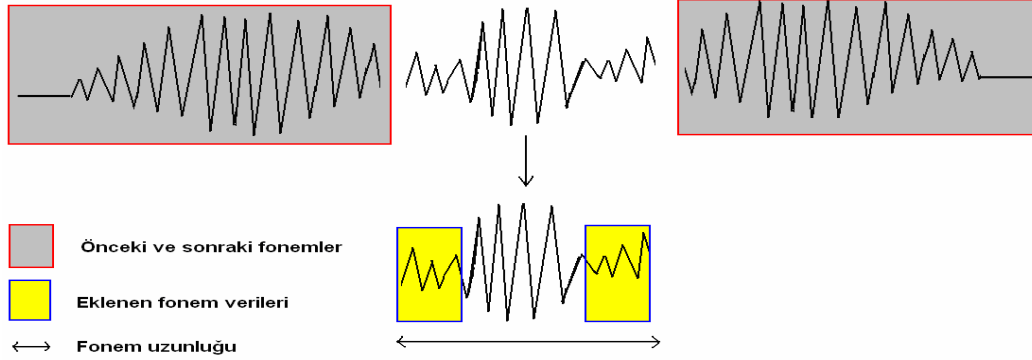
OVY'de vektör boyutlarının eşit olması gerekir. Bunu sağlamak için değişik yaklaşımlar göz önüne alınabilir. Öznitelik vektör boyutlarının eşitlemek için, en uzun söyleyişe ait öznitelik vektörü boyutuna getirmek için sona sıfır veya rastgele sayı eklenebilir. Ya da öznitelik vektörleri, veri tabanında ortalama uzunluğa karşılık gelen uzunluğa getirilecek şekilde baştan sonda kesme veya başa ve sona sıfır ekleme yapılabilir. Öznitelik vektör boyutlarının eşitlemede kullanılacak yöntemler şekil 3.2'de gösterilmektedir.



Şekil 3.2 Fonem öznitelik vektör uzunluklarını eşitleme yöntemleri: a) Baştan ve sondan kesme, b) sıfır ekleme ve c) rastgele sayı ekleme

Deneyisel çalışmalardan elde edilen sonuçlar, sona ve başa sıfır ya da rasgele sayı eklemenin başarımı düşürdüğü görülmüştür. Bu nedenle, kısa fonemlerin öncesi ve sonrasındaki ses verileri alınarak fonemler eşit uzunluğa getirilmiştir.

Bu şekilde fonem uzunluk eşitlemekle tanıma başarımının arttığı gözlenmiştir (Şekil 3.3).



Şekil 3.3 Kısa fonemler için öznitelik vektörlerinin boyunun eşitlenmesi

3.3 Parametrelerin Elde Edilmesi

3.3.1 Çerçeveleme

Ses verisinin karakteristik özellikleri küçük zaman aralıklarında kararlı olmaktadır. Bu yüzden genellikle çerçeve aralığı 20-30 ms olarak alınmaktadır. Her çerçeve kendinden önceki çerçeveyi örtecek şekilde uygulanarak spektral bilgi kaybının önlenmesi sağlanmaktadır (Demirci, M.D. ve Öztaş, O., 2005).

3.3.2 Pencereleme

Ses verisindeki her çerçeveye uygulanarak çerçeve başı ve sonundaki süreksizlikler ortadan kaldırılarak spektral bozulmanın engellenmesi sağlanır. Genel olarak Hamming pencereleme metodu kullanılmaktadır (Rabiner, L., R and Schafer, R., W., 1978).

$$h(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (3.1)$$

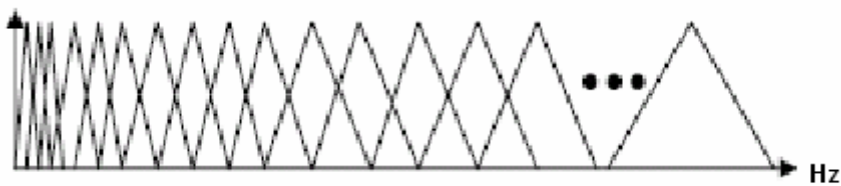
$$y(n) = x(n).h(n) \quad 0 \leq n \leq N-1 \quad (3.2)$$

Eşitlik (3.2)'de $x(n)$, bir çerçevedeki ses sinyalini; $y(n)$, pencerelenmiş sinyali; ve N , çerçevedeki örnek sayısını göstermektedir.

Ses verisinin spektrum özelliğini en iyi şekilde temsil edecek parametreleri bulmak için uygulanan pek çok yöntem bulunmaktadır. LPC, MFCC, PLP yöntemleri bu yöntemlere örnek olarak gösterilebilir. İnsan duyusu ses sinyalinin temel frekanslarını tam olarak doğrusal algılayamaz. İnsan duyusu 1000 Hz'e kadar doğrusal, 1000 Hz'in üstündeki frekansları ise logaritmik olarak algılar. İşte buradan yola çıkılarak çeşitli frekans ölçekleri geliştirilmiştir. En çok kullanılanlardan birisi de mel frekans ölçeğidir.

$$mel(f) = 2595 \log(1 + f / 700) \quad (3.3)$$

Bu ölçek f frekansını mel frekansa dönüştürmektedir. Önvurgulmuş ve Hamming penceresinden geçirilmiş ses sinyali, Mel frekans değerleriyle bağıntılı üçgensel bant geçiren filtre bankasından geçirilir .



Şekil 3.4 Mel bank filtre yapısı

Filtrelerin logaritmik enerji değerleri üzerine ters Fourier dönüşümü (IFT:Inverse Fourier Transform) uygulanarak MFCC'ler hesaplanır. Burada IFT aslında Ayrık Kosinüs Dönüşümü'ne (DCT) karşılık gelmektedir. DCT, MFCC'ler arasındaki ilintiyi de kaldırır. MFCC'ler Eşitlik 3.4'den elde edilir.

$$MFCC_i = \sum_{k=1}^K X_k \cos \left\{ i \left(\frac{k-1}{2} \right) \frac{\pi}{K} \right\} \quad i=1,2,3\dots K \quad (3.4)$$

Burada K , filtre bank sayısını; X_k , k . filtrenin logaritmik enerji değerini; $MFCC_i$; i . Mel frekans cepstral katsayısını göstermektedir.

3.4 Ortak Vektör Yaklaşımı

Bir eğitim setindeki herhangi bir sınıfa ait öznelik vektörleri $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m \in R^n$ ile gösterilsin. Bu sınıf için ortak değışinti matrisi Φ eşitlik (3.5)'den elde edilir.

$$\Phi = \sum_{i=1}^m (\vec{a}_i - \vec{a}_{ort})(\vec{a}_i - \vec{a}_{ort})^T \quad (3.5)$$

Eşitlik (3.6)'da \vec{a}_{ort} , sınıf ortalama vektörünü göstermektedir. OVY, sınıfa ait vektörlerle o sınıfın ortak vektörü arasındaki uzaklığı en küçükleyen bir ölçüt kullanır (Gülmezoğlu, M.B., et al., 2001). Bu en iyileme probleminin çözümü genelleştirilmiş özdeğer probleminin çözümünden elde edilir (Gunal, S., 2007; Gülmezoğlu, M.B., 2007). Buna çözüme göre ortak değışinti matrisinin en küçük özdeğerlerine karşılık gelen özvektörlerin gerdiği uzayda istenilen ölçüt en küçüklenir. Ölçütün en iyilendiği uzaya farksızlık altuzayı (B^\perp) olarak isimlendirilir. Diğer özvektörlerin gerdiği fark uzayıdır (B) ve bu uzay farksızlık uzayına diktir.

Yetersiz veri durumunda ($m < n$) farksızlık altuzayı, $n-m+1$ adet sıfır özdeğere karşılık gelen özvektörler tarafından gerilir (Gülmezoğlu, M.B., et al., 2007). Bu özdeğerler küçükten büyüğe sıralandığında fark ve farksızlık altuzayları için özdeğerler aşağıdaki şekilde seçilir:

$$\underbrace{\lambda_1 < \lambda_2 < \dots < \lambda_{n-m+1}}_{n-m+1} < \underbrace{\lambda_{n-m+2} < \dots < \lambda_{n-1} < \lambda_n}_{m-1}$$

$$\lambda_i = 0 \quad \lambda_i \neq 0$$

Farksızlık Altuzayı Fark Altuzayı

λ_i özdeğerine karşılık gelen özvektör u_i ile gösterilsin. Bu durumda farksızlık altuzay izdüşüm matrisi aşağıdaki şekilde bulunur.

$$P^\perp = \sum_{i=1}^{n-m+1} u_i u_i^T \quad (3.6)$$

Yetersiz veri durumunda, ortak vektör, sınıfın herhangi bir öznitelik vektörünün farksızlık altuzayına izdüşümünden elde edilir.

$$\vec{a}_{ortak} = P^\perp \vec{a}_i \quad i = 1, 2, \dots, m \quad (3.7)$$

Yeterli veri durumunda ($m \geq n$), fark ve farksızlık alt uzayları tahmini olarak belirlenebilir. Ortak değışinti matrisinin sıfırdan farklı n adet özdeğeri olur. Bu özdeğerler küçükten büyüğe sıralandığında ilk k adet küçük özdeğere karşılık gelen özvektörler farksızlık altuzayını gerer. Yeterli veri durumunda ortak vektör, sınıf ortalama vektörünün farksızlık altuzayına izdüşümünden elde edilir:

$$\vec{a}_{ortak} = \sum_{i=1}^k (\vec{a}_{ort}^T u_i) u_i \quad (3.8)$$

Burada \vec{a}_{ort} , öznitelik vektörlerinin ortalamasını göstermektedir:

$$\vec{a}_{ort} = \frac{1}{m} \sum_{i=1}^m \vec{a}_i \quad (3.9)$$

Farksızlık alt uzayını geren en küçük k özdeğer sayısı, özdeğerlerin toplamının belli L yüzdesine karşılık gelecek şekilde seçilebilir (Gülmezoğlu, M. B. et al., 2007).

$$L = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \quad (3.10)$$

Bilinmeyen bir x sesine ait öznitelik vektörünü \vec{a}_x ile gösterilsin. Bu sesin hangi sınıfa ait olduğu belirlenirken önce, \vec{a}_x vektörünün her sınıfın farksızlık altuzayındaki izdüşümünün o sınıfa ait ortak vektörle olan öklit uzaklığı bulunur. Bilinmeyen ses en küçük uzaklığı veren sınıfa atanır.

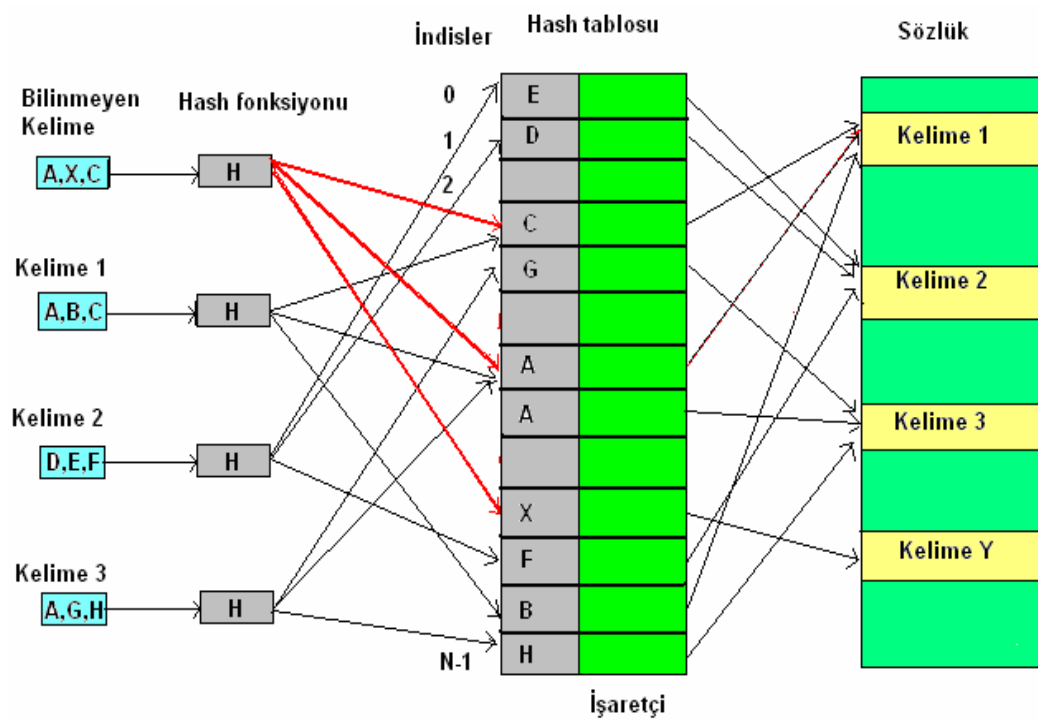
$$c^* = \arg \min_{1 \leq i \leq C} \left\{ \left\| P_i^\perp \vec{a}_x - \vec{a}_{i,ortak} \right\| \right\} \quad (3.11)$$

Eşitlik (3.11)'de C toplam sınıf sayısını göstermektedir.

BÖLÜM 4

REDUNDANT HASH ADRESLEME

RHA adreslemenin temel amacı, sıralı fonem gruplarına karşılık gelen metin birimlerini kullanarak kelime çözümleme işlemini gerçekleştirmektir (Reuhkal, E., et al., 1979). Bütünüyle bir hash algoritması Şekil 4.1’de görüldüğü gibi dört kısımdan oluşmaktadır.



Şekil 4.1 RHA metodunun gösterimi

- Hash fonksiyonu
- Hash tablosu
- Çatışmanın çözülmesi
- Sözlükten uygun kelimenin seçimi

4.1 Hash Fonksiyonu

Hash fonksiyonu, veri arama işlemini bir çırpıda yapmak ve aranana doğrudan ulaşmak için kullanılan bir fonksiyondur. Fonksiyona, arama işleminde kullanılacak anahtar değeri (key) girilir ve karşılığında bir tamsayı (indis) değeri elde edilir. (Reuhkal, E., et al., 1979 ; Erickson , J., 2002).

Verilerin tablodaki yerlerini hash fonksiyonu belirler.

$$h(k)=\text{indis} \quad (4.1)$$

4.2 Hash Tablosu

Hash tablosu veriye bir anahtar (key) yardımıyla erişilen dizidir ve veriler, hash tablosunun ilgili indis değerine karşılık gelen satıra yerleştirilir (Reuhkal, E., et al., 1979). Hash tablosu Şekil 4.2’de verilmektedir.

$h(k)$	$T[h(k)]$
0	Veri 1
1	Veri 2
2	Veri 3
3	Veri 4
4	Veri 5
5	Veri 6
⋮	
M-2	Veri M-2
M-1	Veri M-1

Şekil 4.2 Hash tablosu

4.3 Çatışmanın Çözülmesi

Hash fonksiyonu iki farklı anahtar değeri için aynı değeri üretiyorsa çatışma oluşur.

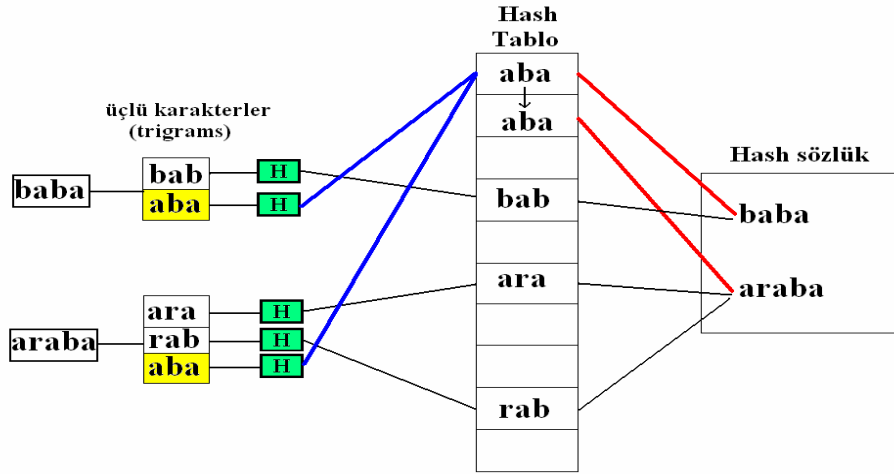
$$h(x) = h(y) \quad (4.2)$$

Çatışmayı en aza indirmek için hash fonksiyonunu iyi seçmek gerekir. Fakat uygulamalarda çatışmayı önleyecek uygun hash fonksiyonu bulunmuyorsa çatışma kaçınılmazdır (Erickson, J., 2002). Çatışmayı önlemek için çeşitli algoritmalar geliştirilmiştir. Bunlardan en çok kullanılanları açık adresleme ve bağlı liste kullanımıdır (Yağmur, O.,2007).

4.3.1 Açık adresleme

Çatışma durumunda veri, çatışmanın olduğu indisten sonraki boş kısma (slot) yerleştirilir. Bu işlem için tablo boyutu kadar farklı değerler üreten hash fonksiyonları $(h_0, h_1, h_2, \dots, h_{M-1})$ oluşturulur. Bu fonksiyonlar çatışmanın olduğu indisten itibaren indis değeri üretirler. Fonksiyonlar kullanılarak boş bir kısım bulunana kadar arama yapılır. Bu şekilde çatışma oluşturan veriler tablodaki boş kısımlara yerleştirilir (Erickson, J., 2002). M , hash tablo boyutunu göstermektedir.

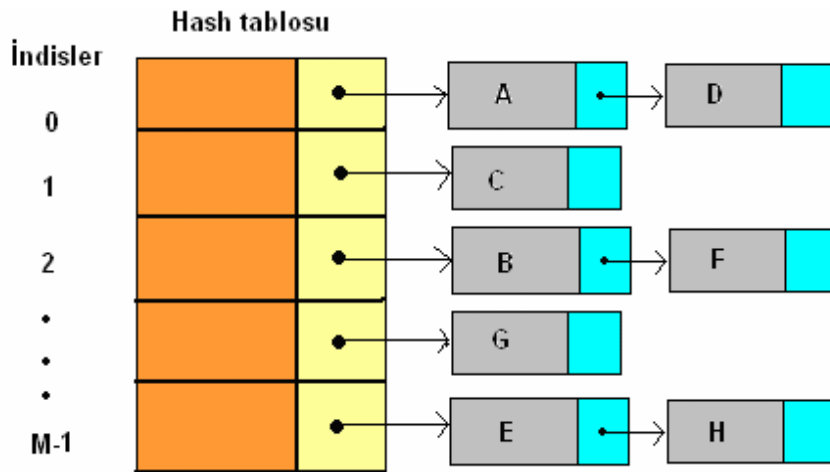
Şekil 4.3'te açık adreslemenin nasıl gerçekleştirildiğine dair örnek görülmektedir. Baba ve araba kelimelerine ait 'aba' üçlü karakter dizisi için çatışma oluşmaktadır. Bu duruda oluşturulan hash fonksiyonları ile çatışmanın olduğu indisten itibaren boş kısım bulunur ve 'aba' üçlü karakteri buraya yerleştirilir.



Şekil 4.3 Açık adresleme örneği

4.3.2 Bağlı liste kullanımı

Çatışma durumunda, çatışan veriler bir liste halinde bağlanırlar (Çölkesen, R., 2004). Hash fonksiyonu bu listelerden birine ait indis ürettiğinde bu bağlı liste üzerinden dolaşarak karar kuralı temelinde doğru kelimeye ulaşmaya çalışır.



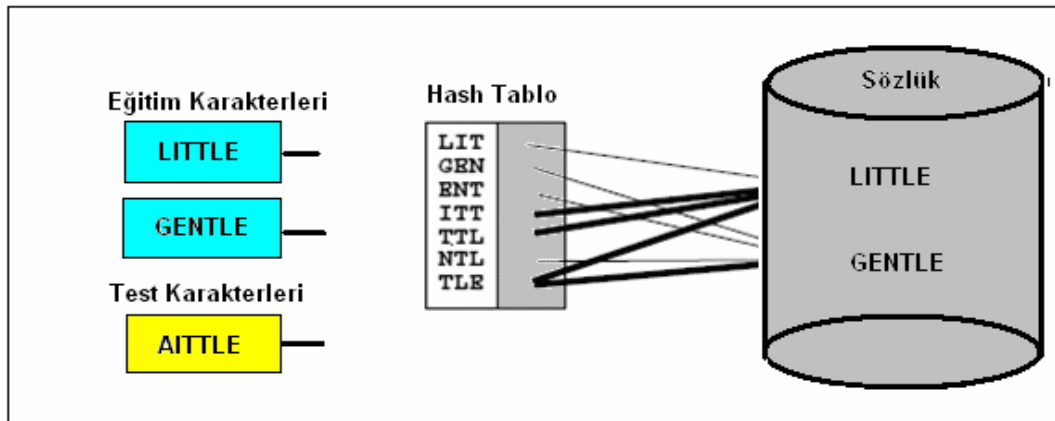
Şekil 4.4 Bağlı liste ile çatışmanın aşılması

4.4 Hash sözlükten Uygun Kelimenin Seçimi

Hash sözlüğün, tanınması amaçlanan tüm kelimeleri kapsamı gerekmektedir. Bilinmeyen bir kelime için hash fonksiyonu ile indis değeri üretilmektedir. İndis değerinin gösterdiği bellek gözündeki adres bilgisi bir işaretçide (pointer) tutulur (Reuhkal, E., et al., 1979; Kohonen, T., et al., 1980; Çölkesen, R.,2004). İşaretçi hash tablosundaki N-gram'lara karşılık gelen sözlük kelimelerini gösterir. Bu sayede bilinmeyen kelime için sözlük kısmında kelime adayları bulunur. Adaylar arasından seçim yapabilmek için aşağıda verilen eşitlik kullanılır.

$$FD(A, B) = \max(n_A, n_B) - n_E \quad (4.3)$$

Burada FD özellik uzaklığı (feature distance) olup, bilinmeyen kelimeye ait N-gramlar ile hash tablodaki N-gramlar arasındaki uzaklığı ölçütünü göstermektedir. Eşitlik (4.3)'deki n_A ve n_B , A ve B dizilerindeki N-gram'ların sayısını; ve n_E ise iki dizi arasındaki benzer N-gram sayısını belirtmektedir. N-gram, sözcüklere ait karakter dizisinin N uzunluklu alt dizisi şeklinde tanımlanır. N-gram'ın boyutu bir ise unigram, iki ise bigram, üç ise trigram ve üçden fazla ise N-gram diye adlandırılmaktadır. Örnek vermek gerekirse 'good' kelimesi için trigram'lar; 'goo' ve 'ood' olmak üzere iki adet, unigram için 'g', 'o', 'o', 'd' olmak üzere dört adettir.



Şekil 4.5 Hash adresleme

Burada hash tablo, eğitim kümesinden elde edilen üçlü karakter dizilerinden oluşturulmuştur. Sözlük kısmı, N-gram tabloyu oluşturan bu üçlülerin indekslerinin (pointers'lar aracılığıyla) bir ya da birkaçının birleştirilmesiyle oluşturulmuş kelimelerden meydana gelmektedir. Görüldüğü gibi sözlük LITTLE ve GENTLE sözcüklerinden oluşturulmuş, test sözcüğü ise AITTLE olarak verilmiştir. Test sözcüğü trigram'lara ayrılacak ve sözlük kısmındaki kelimelere ait N-gramlar kullanılarak özellik uzunlukları bulunacaktır. Bu örnek için FD uzunluklarını bulursak;

$$FD(A_L, B_A) = \max(n_{A_L}, n_{B_A}) - n_{E(L\&A)}$$

$$n_{A_L} = 4, n_{B_A} = 4, n_{E(L\&A)} = 3$$

$$FD(A_{LITTLE}, B_{AITTLE}) = 4 - 3 = 1$$

$$FD(A_G, B_A) = \max(n_{A_G}, n_{B_A}) - n_{E(G\&A)}$$

$$n_{A_G} = 4, n_{B_A} = 4, n_{E(G\&A)} = 1$$

$$FD(A_{GENTLE}, B_{AITTLE}) = 4 - 1 = 3$$

Buradan da anlaşıldığı gibi AITTLE kelimesi, en küçük değeri veren LITTLE kelimesine, GENTLE kelimesinden daha yakındır. Amaç A ve B dizileri arasındaki minimum değeri veren FD değerini elde etmektir. Böylece en küçük değeri veren dizi sınıfına atama yapılarak, kelime çözümlemesi gerçekleştirilmiş olur. Görüldüğü üzere RHA'nın en önemli özelliği, yanlış bulunan kelimelerin hash fonksiyonunun bulacağı indisler yardımı ile sözlük kısmındaki aday kelimelerden birine muhtemelen doğru atamayı yapmayı sağlamasıdır.

BÖLÜM 5

DENEYSEL ÇALIŞMALAR

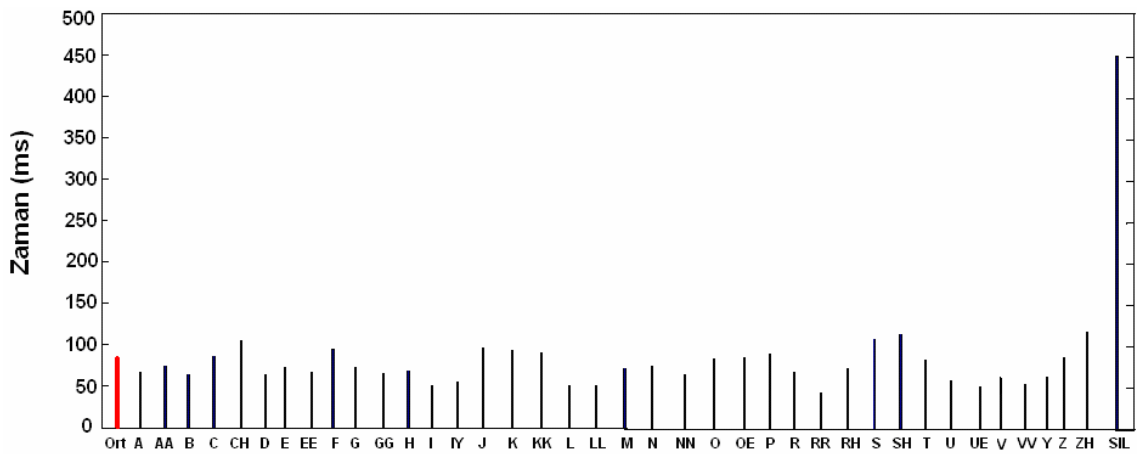
Kelime tanıma oranlarının istenilen düzeyde çıkabilmesi için eğitim seti üzerinden en yüksek başarıyı veren ortak vektörlerin tespit edilmesi gerekmektedir. Bu yüzden kullanılacak parametre ve pencere uzunluğunun en iyi biçimde bulunması hedeflenmiştir. Tez kapsamında kişi bağımlı ve kişi bağımsız olmak üzere iki sınıfta kelime tanıma gerçekleştirilmiştir. Kişi bağımlı tanıma için veri tabanından en fazla fonem sınıfını seslendiren konuşmacının ses verilerinden faydalanılmıştır. Kişi bağımsız tanıma için ise bazı fonem sayılarının azlığından dolayı sınırlamalarla karşılaşmıştır. METU veri tabanındaki fonem uzunlukları ve sayıları Matlab programı ile bulunmuştur. Bu değerlerle; fonemlerin eşitlenecekleri uzunluğun tespitinin yapılması, fonem uzunluğuna eşit uzunlukta kullanılacak pencere boyutunun bulunması ve kişi bağımlı tanıma için gerekli sayıda fonem sınıfı verisinin bulunup bulunmadığı gibi önemli bilgilere erişilmiştir. METU veri tabanındaki fonem sayıları Tablo 5.1’de verilmektedir. Tablo 5.1’de sayıları diğerlerine göre az olan fonemler sarı renkle gösterilmiştir.

A	AA	B	C	CH	D	E	EE	F	G	GG	H	I	IY	J	K	KK	L	LL
2835	20515	5129	1991	2781	8670	5680	13407	1125	2322	485	2194	9804	17700	150	4194	5665	6120	7148
M	N	NN	O	OE	P	R	RR	RH	S	SH	T	U	UE	V	VV	Y	Z	ZH
6776	396	13120	5641	1671	1911	348	10149	4479	6119	3509	7371	6228	4087	284	2168	6963	2404	988

Tablo 5.1 METU veri tabanı fonem sayıları

Eğitim kümesi için OVY ile her fonem sınıfı için eşit sayıda örnek alınacağı düşünüldüğünde, özellikle kişi bağımsız tanıma çalışmalarında bu beş fonem sınıfının çalışmayı sınırladığı görülmüştür. Örneğin her fonem sınıfı için 900 eğitim ve 100 test örneği kullanıldığında, belirtilen altı fonem sınıfının örnek değerleri aşıldığı için sorun oluşmaktadır. Bu sorunu aşabilmek için veri tabanındaki bazı allofonların birleştirilerek tek fonem sınıfı haline gelmesi sağlandı. J ve V fonemleri ise çok az kelime geçmeleri ve veri tabanındaki birçok kişi tarafından seslendirilmemeleri sebebiyle eğitim ve test setinden çıkartılmıştır.

Şekil 5.1’de otuzsekiz fonem sınıfının ortalama süreleri milisaniye olarak verilmektedir. Şekildeki kırmızı çubuk, veri tabanındaki sessiz (SIL) kısımlar alınmadan fonem sınıflarının hesaplanan ortalama fonem süresini göstermektedir. Veri tabanında fonemlerin ortalama uzunluğu yaklaşık 80 ms (1280 örnek) olarak bulunmuştur. Deneysel çalışmalar sonucunda, en iyi tanıma sonuçları bu uzunlukta elde edilmiştir.



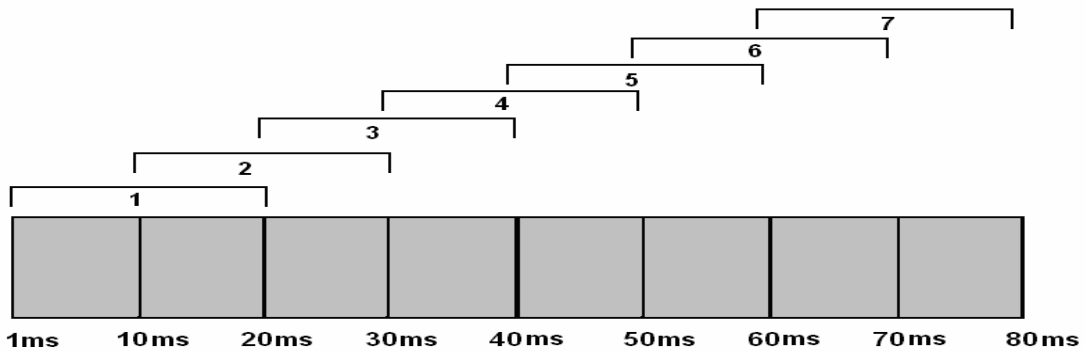
Şekil 5.1 METU veri tabanı ortalama fonem uzunlukları

5.1 Parametre Seçimi

Bir fonemin merkez kısmındaki veriler , baş ve son kısımlarına göre o fonem hakkında daha belirleyici özelliklerin elde edilmesini sağlamaktadır. Ayrıca seslendirilen kelimeleri oluşturan fonem dizilerindeki her bir fonem birbirleri içine girişim (karışım) yapabilmektedir. Bu şekilde fonların birbirine girişim yapmasına ortak eklemleme (coarticulation) etkisi denmektedir. Bu etki tanıma sırasında ortak eklemleme etkisi ile fonem geçişlerinin olduğu kısımlarda nicemleme hatasının oluşmasına yol açmaktadır. Kohonen ve grubu bu etkinin en az şekilde görüleceği fonem merkezlerini referans olarak bir pencere modeli geliştirmişlerdir (Mantysalo, J., et al., 1992). Bu pencere ortak eklemleme etkisinin fazla olduğu fonem geçiş yerlerinde geniş aralıklar, az olduğu fonem merkezlerinde ise dar aralıklar alınarak uygulanmıştır. Pencere uzunluğu ise bir fonem uzunluğuna karşılık gelecek şekilde

oluşturulmuştur. Bu şekilde uygulanacak bir pencere ile ortak eklemelenmenin etkisi düşük düzeyde kalacak ve tanınan fonem dizilerindeki fonem geçiş yerlerinde daha az nicemleme hatası oluşacaktır. Parametre olarak pencerenin ses verisi üzerindeki her konumu için elde edilen bağlam öznitelik vektörleri kullanılmıştır. Pencere yan kısımlarda geniş, orta kısımda ise dar olmak üzere toplam 7 bölüme ayrılmış, her bölümün 10 ms'lik kısmında öznitelik vektörleri hesaplanarak ortalamaları alınmış ve elde edilen öznitelik vektörleri sıralanarak bağlam öznitelik vektörü (context vector) bulunmuştur (Mantysalo, J., et al., 1992).Pencere ses verisi üzerinde 10 ms kaydırılarak uygulanmıştır. Pencerenin her konumu için fonem sınıflama yapılmış ve fonem sınırlarının tespiti için fonem geçiş yerlerindeki nicemleme hatalarından faydalanılmıştır. Bu pencere yapısı tüm fonemi temsil edecek kadar uzun olması ve öznitelik vektörlerinin ortak eklemelenme etkisinin düşük kalacak şekilde elde edilmesini sağlaması bakımından deneysel çalışmada kullanılmış ancak şekil 5.2'de belirtilen pencere yapısı ile karşılaştırıldığında daha kötü sonuçlar verdiği görülerek kullanımından vazgeçilmiştir. Veri tabanı incelendiğinde çoğu fonem uzunluğunun 100 ms civarında ya da altında değerler aldığı görülmüş, bazı fonemlerin ise 30 ile 60 ms uzunluğunda olduğu tespit edilmiştir. Deneysel çalışmada ise 150, 120, 100, 80, 60 ms'lik pencerelere uzunlukları kullanılarak elde edilen fonem tanıma başarımının düşük kaldığı görülmüştür.

Şekil 5.2'de görülen pencere modeli ve uzunluğu kullanımının diğer pencere tiplerine göre daha başarılı sonuçlar verdiği gözlenmiştir. Elde edilen bu sonuçlardan yola çıkarak çalışmalara 80 ms'lik pencere uzunluğu ile devam edilmiştir.



Şekil 5.2 Çalışmada kullanılan pencere yapısı

Deneysel çalışmalarda fonemi temsil eden öznitelik vektörü 80 ms'lik süre üzerinden elde edilmiştir. Burada 20 ms çerçeve üzerinde elde edilen öznitelik parametreleri arka arkaya eklenerek öznitelik vektörü oluşturulmaktadır. Öznitelik çıkartılırken çerçeveler arası 10 ms üst üste bindirme yapılmıştır. Bu durumda 80 ms için 7 çerçeve üzerinde parametre çıkartımı yapılmaktadır. 20 ms çerçeve üzerinden 12 MFCC parametresi ve çerçeve enerjisi elde edilmektedir. Deneysel çalışmada çerçeve enerjisi yok ve var durumu için 84 (12*7) ve 91 (13*7) boyutlu öznitelik vektörleri ayrı ayrı kullanılarak başarımlar elde edilmiştir.

5.2 Fonem Sınıflama Başarımları

En iyi başarımları verecek fonem sınıf modellerini bulabilmek için test kümesi üzerinden çalışmalar yapılmıştır. Kişi bağımlı tanımada veri tabanındaki bir erkek konuşmacı kullanılırken kişi bağımsız tanımada 100 konuşmacının ses verilerinden faydalanılmıştır. Ardından kişi bağımlı ve kişi bağımsız eğitim-test setleri için fonem öznitelik vektörleri elde edilerek her fonem sınıfı için en yüksek başarımları veren izdüşüm matrisleri bulunmuş ve tanımada kullanılmak üzere kaydedilmiştir. Kişi bağımsız kelime tanımada yeterli veri durumu oluşmaktadır. Kişi bağımsız tanımada her fonem sınıfının 300 örneği eğitim, 100 örneği ise test verisi olarak kullanılmıştır.

Test kümesi üzerinden en yüksek başarımları veren ünlü ve ünsüz fonem sınıflarına ait yaklaşık ortak vektörler bulunurken, küçükten büyüğe doğru sıralı 91 adet özdeğerden ilk 86 tanesi ünlü, 85 tanesi ise ünsüz fonem sınıfları için kullanılmıştır. Bu durumda çalışmada en iyi fonem ve kelime sınıflama yapan modeller için toplam enerjinin %23.9'u ve %17.6'sı sırasıyla ünlü ve ünsüz fonem sınıfları için alınmıştır.

$$L_{\text{ünlü}} = \frac{\sum_{i=1}^{86} \lambda_i}{\sum_{i=1}^{91} \lambda_i} = 0.239$$

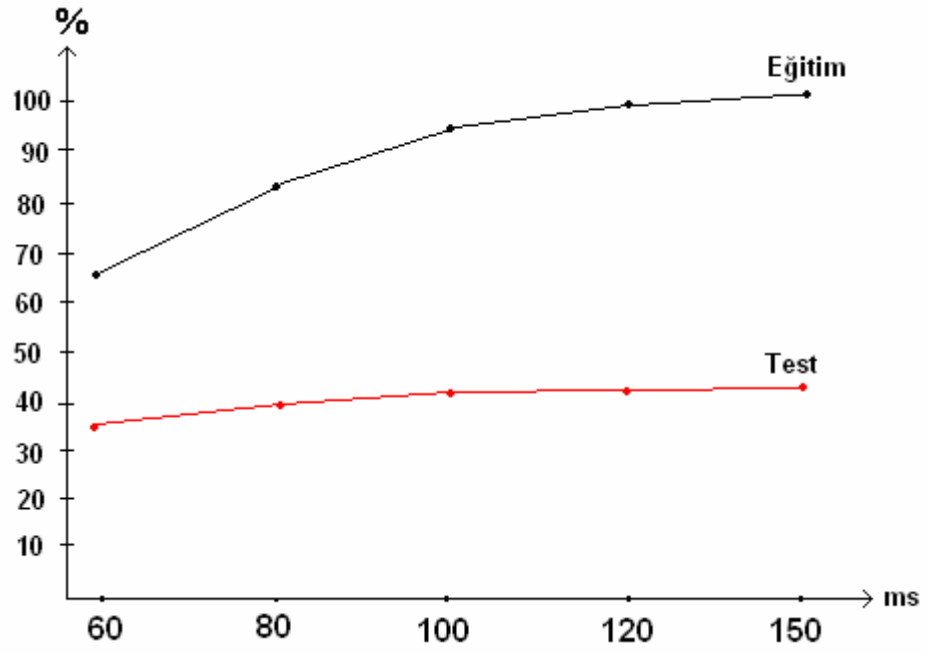
$$L_{\text{ünsüz}} = \frac{\sum_{i=1}^{85} \lambda_i}{\sum_{i=1}^{91} \lambda_i} = 0.176$$

Beş farklı fonem uzunluğu için elde edilen başarımlar Şekil 5.3 ve 5.4 'te verilmektedir. Başarımlar için oluşturulan tablolar Ek-E ve Ek-F'da verilmiştir. Ek-E'te 10 ünlü, 22 ünsüz fonem sınıfı ve Ek-F'da 8 ünlü, 19 ünsüz fonem sınıfı için başarımlar elde edilmiştir. Tablolardaki isimler ve kısaltmalar aşağıda verilen anlamları taşımaktadır.

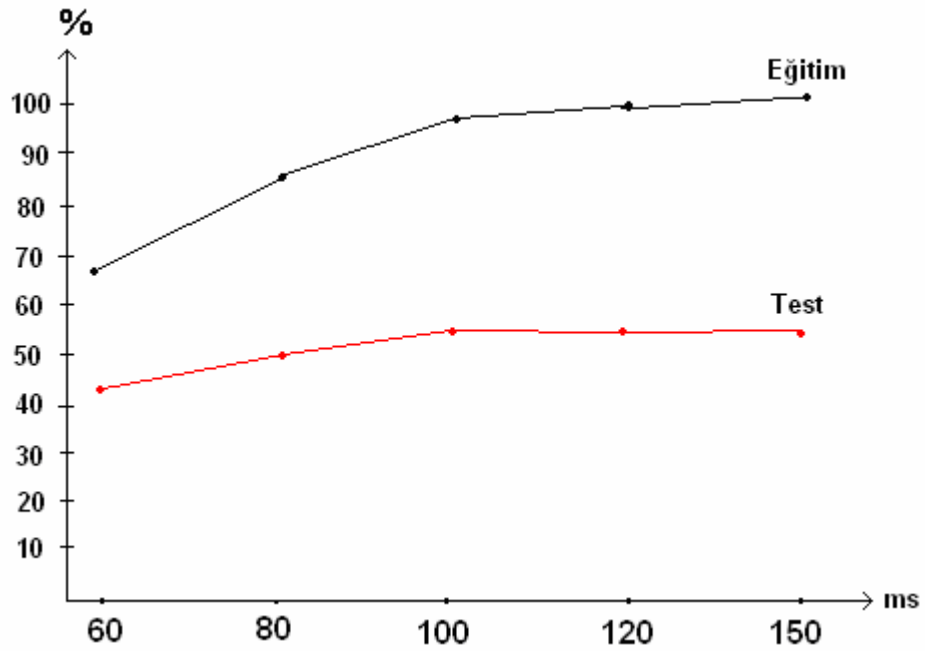
- Eğitim-Test (m)* : Eğitim ve test setlerinde kullanılan enerji+MFCC parametre sayısı.
 Eğitim-Test (m) : Eğitim ve test setlerinde kullanılan MFCC parametre sayısı.
 U-UZ : Ünlü ve ünsüz fonem sınıfları.
 KB-KBZ : Kişi bağımlı ve kişi bağımsızlık.

Ünlü fonemler : AA, A, E, EE, O, OE, U, UE, I, IY

Ünsüz fonemler : B, C, CH, D, F, G, H, K, KK, L, LL, M, NN, P, RH, RR, S, SH,
 T, VV, Y, Z



Şekil 5.3 10 ünlü fonem sınıfı için farklı sürelerde elde edilen başarımlar



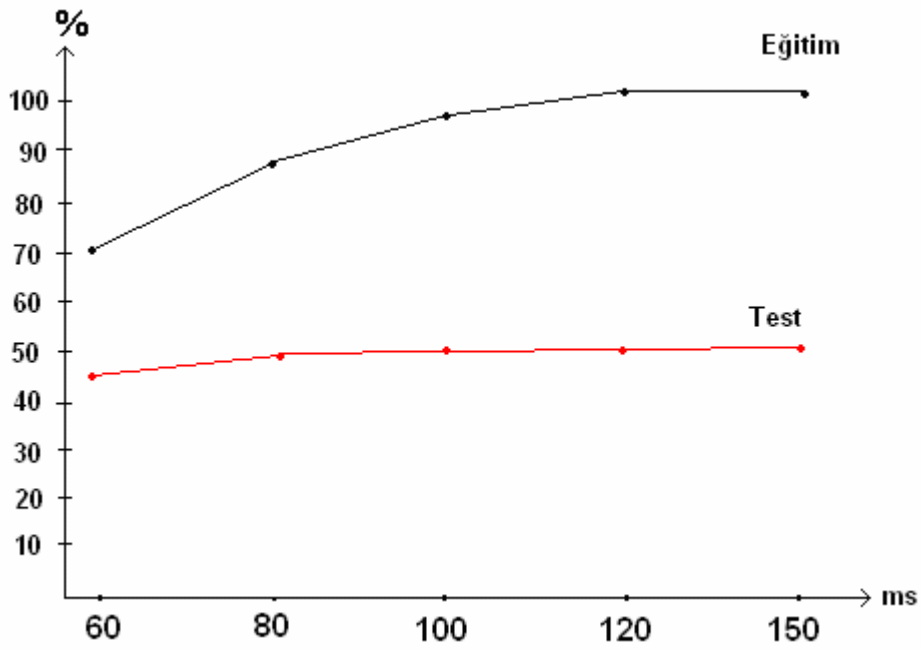
Şekil 5.4 22 ünsüz fonem sınıfı için farklı sürelerde elde edilen başarımlar

Yukarıda elde edilen tanıma oranlarının düşük olması allofonlardan kaynaklanmaktadır. (K, KK, L, LL vb... fonemlerine karşılık gelen fonlar). Bu fon sınıfları diğer fonlara göre daha fazla ortak özellik içerdiklerinden birbirlerinin

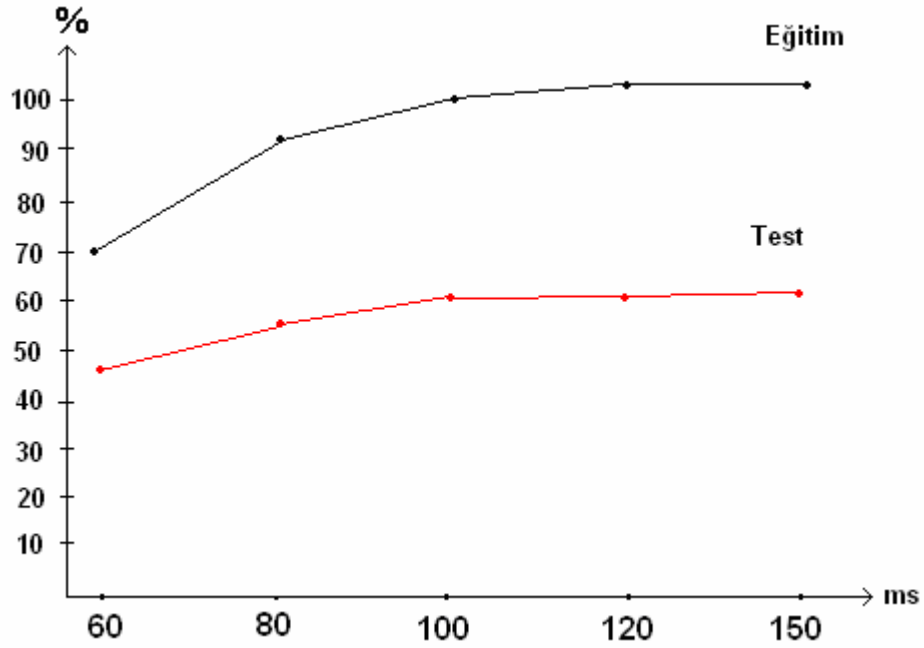
sınıflarına atanarak tanıma oranının düşük kalmasına sebep oluyorlar. Fakat kelime tanıma aşamasında daha yüksek başarımlar elde edilmesine katkıda bulunuyorlar. Eğer tüm allofonların birleştirilmesi ile elde edilen tanıma oranlarını aşağıdaki tablolardan incelersek gerçekte tanıma oranlarının ne kadar olduğunu görebiliriz.

Ünlü fonemler : A, E, O, OE, U, UE, I, IY

Ünsüz fonemler : B, C, CH, D, F, G, H, K, L, M, N, P, R, S, SH, T, VV, Y, Z



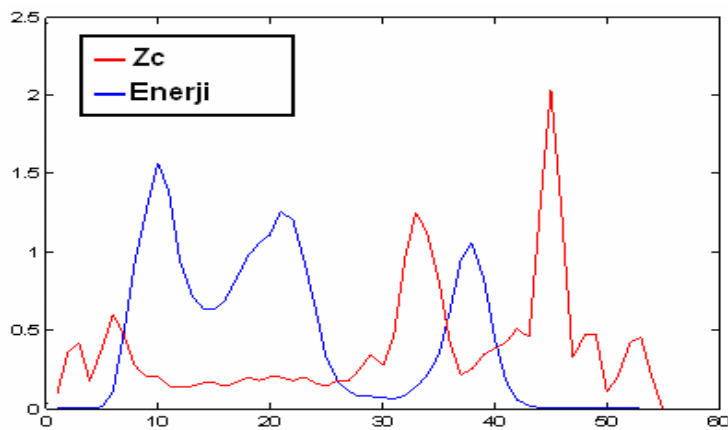
Şekil 5.5 8 ünlü fonem için farklı sürelerde elde edilen başarımlar.



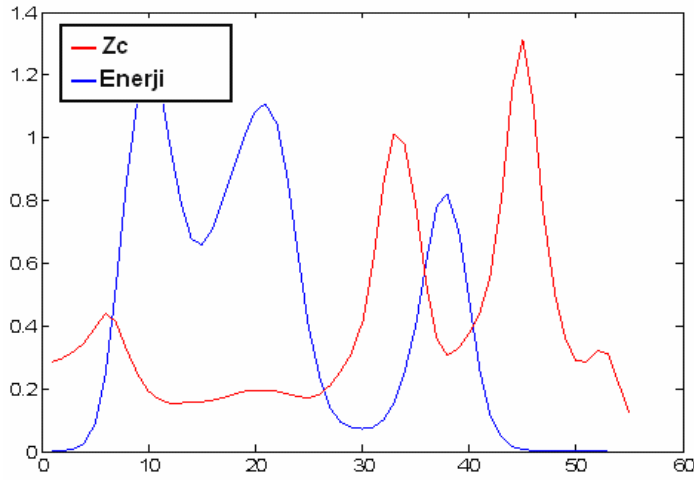
Şekil 5.6 19 ünsüz fonem için farklı sürelerde elde edilen başarımlar

5.3 Kelime Çıkartma

Yalıtık ses verisinin fonem sınırlarının tespiti için enerji ve sıfır geçiş oranlarından faydalanılmıştır. Sabit bir enerji eşik değeri bulunarak ses verisi ünlü ve ünsüz kısımlara ayrılmıştır. Bu işlem yapılmadan önce bulunan enerji ve sıfır geçiş değerleri düzgünleştirilmiştir (smoothing). Bununla ünlü ve ünsüz fonem sınırları bulunurken yanlış kısımların alınmaması amaçlanmıştır.

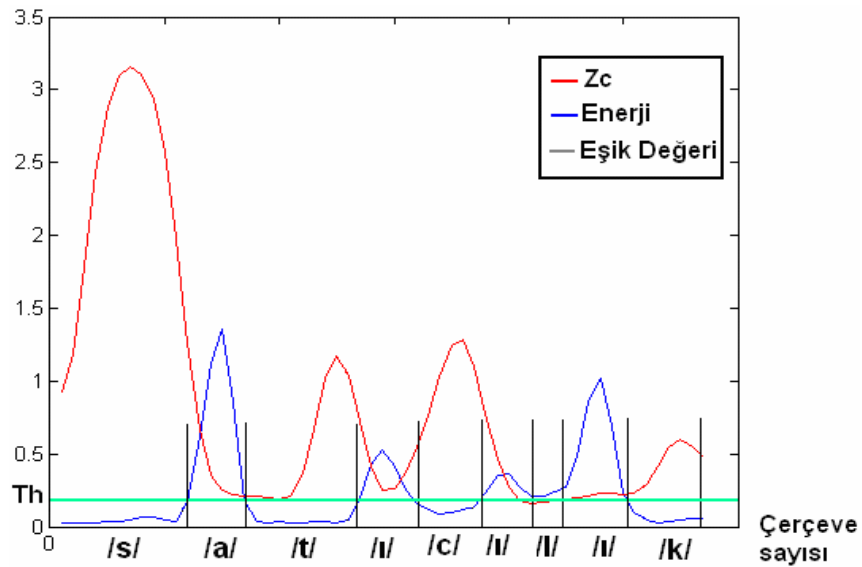


Şekil 5.7 'Gülücük' kelime verisi için enerji ve sıfır geçiş oranı



Şekil 5.8 'Gülücük' kelime verisi için düzgünleştirilmiş enerji ve sıfır geçiş oranı

Bulunan eşik değeri ile ünlü ve ünsüz fonem sınırları kısmen tespit edilebilmiştir. Çünkü bazı ünsüz fonemler iki ünlü fonem arasında yüksek enerji değerine ve düşük sıfır geçiş oranına sahip olabilmektedir. Bu durumu çözmek için yapılan çalışma, aşağıda örnek olarak verilen 'satıcılık' kelimesi için açıklanmıştır.



Şekil 5.9 'Satıcılık' kelimesi için eşik değeri ile fonem sınırlarının tespiti

Ses verisinin enerji eşik değerinden büyük kısımlar ünlü olarak, küçük değerli kısımlar ise ünsüz olarak atandı. Şekilde de görüldüğü gibi /l/ ünsüz fonemi iki /l/ fonemi arasında daha düşük enerji değerine sahip fakat ünlü sınıfına atanıyor. Bu sorunu çözmek için ses verisinin ünlü kısımlarındaki çatalı kısımların kaç çerçeveden oluştuğuna bakıldı. Bu kısımları oluşturan çerçeve sayısının yeterince büyük olması durumunda, yerel minimum noktasının ileri ve gerisindeki toplam 6 çerçeve alınarak ünsüz sınıfına atama yapılmıştır. Bu yöntemin dışında eşik değeri kullanmadan, sadece pencere kaydırma yöntemi ile fonem sınıflama işlemi karşılaştırıldığında eşik değeri kullanılan metodun daha başarılı sonuçlar verdiği görülmüştür. Ancak ortak eklemleme etkisinden dolayı fonemler birbirleri üzerine girişim yapabilmekte, bu yüzden eşik değeri ile bulunan fonem sınırları eksik olarak alınabilmekteydi. Bununla birlikte iki metodun da birbirlerine göre üstün oldukları durumlar gözlenmiştir. Bu yüzden iki metodun birlikte kullanılmasıyla daha iyi fonem sınıflama başarımları elde edileceği düşünülmüştür. Kullanılan yöntemler için isimlendirme yapılmıştır.

Yöntem 1’de tüm fonem sınıfları birlikte kullanılarak oluşturulan modeller üzerinden eşik değeri kullanmadan, sadece pencere kaydırılarak fonem sınıflama yapılmıştır. Yöntem 2A’da ise ayrı ayrı oluşturulmuş ünlü ve ünsüz fonem modelleri için eşik değeri ile bulunan ünlü ve ünsüz sınırlarını kullanarak kelime çıkartılmaktadır. Yöntem 2B’de ayrı ayrı oluşturulmuş ünlü ve ünsüz fonem modelleri için eşik değeri kullanmadan, sadece pencere kaydırılarak fonem sınıflama yapılır. Yöntem 2A ve 2B birlikte kullanılarak eşik değerinden dolayı bulunamayan ünsüz ve ünlü fonemlerin bulunması sağlanmaktadır.

Eşik değeri kullanmadan fonem sınıflama metodunda tüm sözcük verisi, ünlü ve ünsüz fonem sınıflarına göre ayrı ayrı sınıflandırılmıştır. Ayrı sınıflandırma yapmanın, (ünlü ve ünsüz) tüm fonem sınıflarını kullanarak sınıflama yapmaktan daha verimli sonuçlar verdiği gözlenmiştir. Yöntem 1 ile yapılan çalışmada aşağıda verilen örnek sözcükler için belirtilen fonem dizileri elde edilmiştir.

‘Siyah’ sözcüğü için;

F , D , D, T, KK, KK, S, S, S, S, S, S, S, F, F, F, IY, IY, IY, IY, IY, IY, SH
SH, SH, SH, SH, SH, SH, SH, SH, SH, SH, SH, SH, SH, NN, SH, IY

‘Hayat’ sözcüğü için;

F,D,D,T,KK,KK,KK,L,H,H,SH,CH,UE,AA,AA,EE,IY,IY,IY,SH,SH,SH,SH,
SH, SH, SH, SH, RH, RH, SH, SH, SH, SH

Görüldüğü gibi ‘siyah’ sözcüğü için /S/ ve /IY/ fonemleri bulunurken /Y/, /A/ ve /H/ fonemleri bulunamamıştır. Benzer şekilde ‘hayat’ sözcüğü için /H/ ve /AA/ fonemleri bulunurken /Y/, /A/ ve /T/ fonemleri bulunamamıştır. Bu yönteminde; ses verisini oluşturan fonemler tüm fonem sınıfları hesaba katılarak sınıflandırma yapılmaktadır. Yani kaydırılan pencere, ses verisinin ünlü kısımlarından geçerken ünsüz, ünsüz fonem sınıflarından geçerken ise ünlü fonem sınıfları da karar kuralına dahil edildiğinden, özellikle çok kısa uzunluklu fonemlerin tanınmasında hata oranının artmasına yol açılmaktadır.

Aynı sözcükler kullanılarak ünlü ve ünsüz fonem sınıfları için ayrı ayrı sınıflama yapıldığında ise ‘siyah’ sözcüğü için Yöntem 2B ile ünsüz fonem sınıflamada aşağıdaki fonem dizisi bulunmuştur:

T, D, G, K, KK, K, S, S, S, S, S, S, F, SH, VV, L, Y, Y, Y, Y, L, L, L, H, H, H, H, H, L, L,
H, H, NN, RH, L, L, F, F.

Aynı sözcük için ünlü fonem sınıflama yapıldığında ise;

U, U, U, U, U, OE, I, I, I, I, I, I, I, I, I, I, IY, E, E, A, A, A, A, A, A, A, A, A, A, A, A,
U, I, I, I, U fonem dizisi ortaya çıkar.

Yöntem 2A ile fonem sınıflamada ise aşağıda belirtilen fonem dizileri elde edilmiştir.

‘Siyah’ (ünlü) : I, I, I, I, A, A, A, A, A, A, A

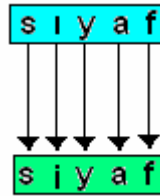
‘Siyah’ (ünsüz) : T, D, G , K, KK, K, S, S, S, S, S, S, Y, Y, Y, Y, L, L, L, H, H, NN, RH,
L, L, F, F

‘Hayat’ (ünlü) : AA, AA, AA, A, A, A, A, A, A, A, AA

‘Hayat’ (ünsüz) : T, D, G, KK, KK, K, KK, P, L, NN, RH, RH, NN, L, L, L, L, L, VV, VV, NN, NN, NN, H

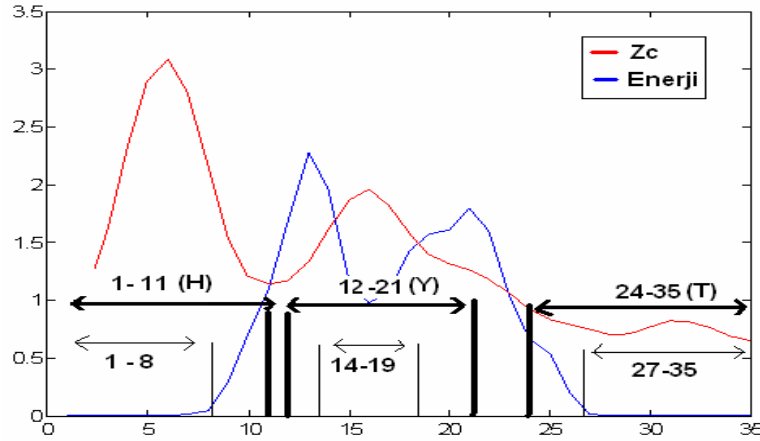
Görüldüğü gibi ‘siyah’ kelimesi için ünlü fonem dizisinde /Y/ fonemi bulunmamaktadır. Bu durumda elde edilen karakter dizisi ‘sıyaf’ şeklindedir. ‘Hayat’ kelimesi için ünsüz fonem dizisinde /H/ fonemi bulunmamaktadır ve ‘kalah’ karakter dizisi elde edilmiştir. OVY karar kuralı bulunan fonem sınırları içinde uygulanmaktadır.

Elde edilen bu sonuçlardan yola çıkarak öncelikle eşik değeri kullanarak fonem sınıflama metodu ile (Yöntem 2A) karakter dizileri bulunmuştur. (‘sıyaf’, ‘kalah’ vb...). Ardından Yöntem 2A ve 2B ile o sözcük için ünlü ve ünsüz fonem sınıflarına karşılık gelen karakter dizileri bulunarak Yöntem 2A ile bulunan karakterlerin yerine bu karakterler sırası ile yerleştirilmiştir. (‘siyah’ için ‘s, y, f’ ve ‘i, a’, ‘hayat’ için ‘h, n, n’ ve ‘a, a’). Böylece Yöntem 2A ve 2B ile bulunan ‘syf’ ve ‘ia’ karakterleri sırasıyla ‘sıyaf’ karakterlerinin yerine kullanılmıştır. Sonuç olarak ‘sıyaf’ karakter dizisi bulunmuştur.



Şekil 5.10 ‘sıyaf’ karakter dizisinin iyileştirilmesi

Bu karakter dizileri bulunurken Yöntem 2A ile bulunan fonem sınırlarından faydalanılmıştır. Örneğin ‘hayat’ sözcüğü 35 çerçeveden oluşmaktadır. İlk 8 çerçeve /H/ fonemi, 9-13 arası /A/, 14-20 arası /Y/, 21-26 arası /A/ ve 27-35 arası /T/ fonemini bulmak için atanmıştır. Ardından Yöntem-2’deki ünlü ve ünsüz fonem dizileri için çerçeve aralıklarının ortasındaki değerlerin belli sayıda ileri ve gerisindeki çerçevelerde dahil edilerek bulunan yeni fonem sınırları için en küçük uzaklığı veren fonem sınıfı seçilmiştir. Ünsüz fonem sınırları için bu durum Şekil 5.11 ile belirtilmiştir.



Şekil 5.11 'Hayat' kelimesi için ünsüz fonem sınırlarının geniş alınması

Yapılan incelemelerde bazı fonemler için bulunan aralığın en fazla 15-25 çerçeve uzunluğunda, en az ise 3-5 çerçeve uzunluğunda olabildiği görülmüştür. Fonemlere ait bilgilerin büyük kısmının fonemlerin merkez bölgelerinde bulunduğundan yola çıkarak (Mantysalo, J., et al., 1992), yaklaşık 20 çerçevelik uzun aralıklı kısımlar için tam orta noktanın (10. çerçeve) belli miktar ileri ve gerisindeki (örneğin 5-15 arası) çerçeveler için sınıflama yapılırken; kısa aralıklı kısımlar için (3-5 arası), bir miktar genişletme (örneğin 8 çerçeve) yapılarak sınıflama gerçekleştirilmiştir. Burada sınır genişletmedeki amaç; ortak eklemlemeden kaynaklı olarak, Yöntem 2A ile bulunamayan fonem sınıflarını karar kuralına dahil etmektir. Tanınması hedeflenen kelimelerin tümüne bu yöntem uygulandığında pek çok kelimeye ait karakter dizilerinin düzeldiği görülürken, az sayıdaki kelimeler için ya düzeltme görülmemiş ya da daha kötü karakter dizileri elde edilmiştir.

Sadece Yöntem 2A ile yapılan kişi bağımsız kelime tanıma çalışmasında %71.11, kişi bağımlı tanımada ise % 80 oranında başarımlı sağlanmıştır. Karakter iyileştirme yöntemi ile kişi bağımsız kelime tanıma oranı %72.22, kişi bağımlı kelime tanımada ise % 82.5 oranına erişilmiştir. Kişi bağımlı ve kişi bağımsız tanımada kullanılan kelimeler, kelimelere karşılık bulunan karakter dizileri ve tanınan kelimeler Ek-A , Ek-B, Ek-C ve Ek-D'de verilmiştir.

Aşağıdaki tabloda bazı örnek kelimeler için verilen yöntemler kullanılarak çıkartılan kelimeler verilmektedir.

Kelime	Yöntem 2A	Yöntem 2A ve 2B
hayat	kalah	hanan
şehir	seyer	şeyer
tamam	kayan	vamam
sırada	salapö	sırodö
çocuk	çöcöp	çöcök
kısa	tasar	kısar
bizi	vızı	bizi
buhar	pıhar	bohar
kanun	karını	kanını
şekil	şukil	şekil
beni	denir	beniy
durumu	dulumul	durumul
fakat	fikak	farat
dolayı	vololir	doloyir
tavus	kavof	tavof

Tablo 5.2 80 ms pencere uzunluğu ile kelime çıkartma

Yukarıdaki tabloda bulunan 6 kelime alınarak 60, 100, 120 ve 150 ms'lik pencere uzunlukları ile kelimelere ait karakter dizileri aşağıdaki tabloda verilmiştir. Bulunan kelimeler incelendiğinde 80 ms'lik pencere uzunluğunun daha başarılı sonuçlar verdiği görülebilir.

Kelime	Yöntem 2A	Yöntem 2A ve 2B
kısa	tafat	tanat
siyah	seyef	seyef
tamam	tavek	tavek
şehir	şücek	şacek
sırada	firöfi	foröfi
hayat	hanaf	hanaf

Tablo 5.3 60 ms pencere uzunluğu ile kelime çıkartma

Kelime	Yöntem 2A	Yöntem 2A ve 2B
kısa	tösar	çesar
siyah	sıvır	sivir
tamam	tılam	çılam
şehir	şürür	şirür
sırada	sıvolö	süsolö
hayat	çilel	hııl

Tablo 5.4 100 ms pencere uzunluğu ile kelime çıkartma

Kelime	Yöntem 2A	Yöntem 2A ve 2B
kısa	tösar	tasar
siyah	sivan	sivan
tamam	tovöm	tavem
şehir	şiyir	şiyir
sırada	çisohı	sısohı
hayat	tölöv	hahöv

Tablo 5.5 120 ms pencere uzunluğu ile kelime çıkartma

Kelime	Yöntem 2A	Yöntem 2A ve 2B
kısa	tılıl	hısıl
siyah	çicen	sicen
tamam	henan	tıhan
şehir	liliy	şiliv
sırada	lısırı	tısırı
hayat	töver	liver

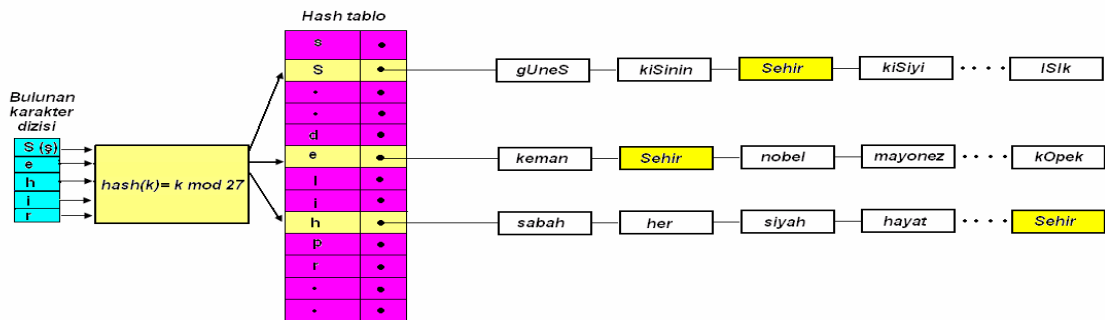
Tablo 5.6 150 ms pencere uzunluğu ile kelime çıkartma

5.4 RHA ile Kelime Çözümleme

Fonemlerin alfabetik karşılıkları aşağıda verilen hash fonksiyonu ile hash tablosuna yerleştirilir:

$$hash(k) = k \text{ mod } 27 \quad (5.1)$$

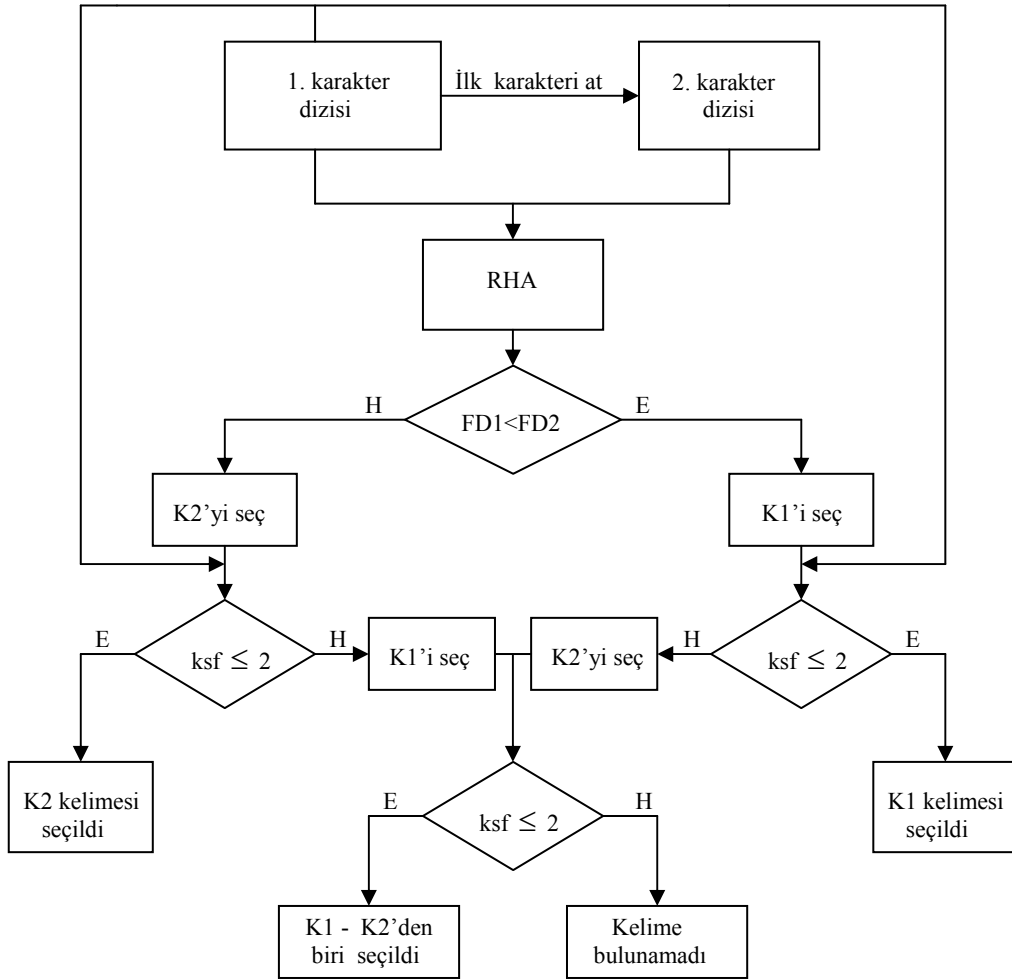
Burada k , fonemin alfabetik karşılığının (harfin) ASCII değeridir. Deneysel çalışmada, veri tabanında bazı fonemler sık geçmediği için alınmamıştır. Bu nedenle toplam 27 foneme karşılık gelen alfabetik harfler hash tablosu belirlenirken kullanılmıştır. Tez kapsamında kelime çıkartmada bulunan kelimelere ait tekli karakter (unigram) dizileri kullanılarak kelime çözümleme yapılmıştır. Çatışmanın çözülmesi için bağlı liste yöntemi kullanılmıştır. Aşağıda verilen Şekil 5.12’de ‘şehir’ kelimesi için örnek kelime çözümleme gösterilmiştir.



Şekil 5.12 ‘Şehir’ kelimesi için bağlı liste ile kelime çözümleme

Çalışmada kelime çıkartma ile bulunan kelimelere ait karakterler için hash fonksiyonu ile indis değeri bulunmuştur. İndis değeri hash tablodaki ilgili karaktere ait bağlı listeyi gösterir. Bu bağlı liste üzerindeki tüm kelimelerle çıkartılan kelime arasındaki özellik uzaklıkları bulunur. Bu işlem kelime çıkartmada bulunan kelimenin tüm karakterlerine uygulanır. Elde edilen tüm özellik uzaklıkları içerisinde en küçük uzaklığı veren hash sözlükteki kelime, tanınan kelime olur.

Fonem tanıma yapıldıktan sonra elde edilen fonem dizileri kullanılarak tanınacak kelimeye ait harf dizileri bulunur. Bu kelime sözlükteki kelimelerden biri olabilir veya sözlükteki kelimelerden birine benzerde olabilir. RHA, sözlükteki en çok benzer kelimeyi bulmakta kullanılır. Kelime tanıma aşamasında bazı harfler yanlış tanınmakta, bazıları ise kelime içinde bulunması gerekirken ortaya çıkmamaktadır. Bu durum sistemin kelime tanıma başarımlarını etkilemektedir. Tez çalışmasında RHA yönteminde bazı iyileştirmeler yapılarak tanıma başarımlarının yükseltilmesi üzerinde çalışılmıştır. Bunun için Şekil 5.13'de belirtilen işlemler yapılmıştır.



Şekil 5.13 RHA ile yapılan iyileştirme algoritması

Kelime çıkartma sonucu bulunan kelimeye ait birinci karakter dizisi (KD1) için en küçük özellik uzaklığını veren (FD1) ilk aday kelime bulunur (K1). Ardından karakter dizisinin (KD1) ilk karakteri atılarak bulunan karakter dizisi (KD2) için yeniden çözümleme yapılır ve ikinci aday kelime bulunur (K2). Ardından en küçük özellik uzaklığı değerini (FD) veren indis ve ona karşılık gelen kelime seçilir. Seçilen kelimeye ait karakter sayısı ile başlangıçta kelime çıkartma ile bulunan karakter dizisi (KD1'in) sayısı karşılaştırılır. Karakter sayısı farkı (ksf) 2'den büyük ise seçilen aday kelime yerine diğer aday kelime seçilir. Bu şekilde seçilen aday kelimenin karakter sayı farkı da 2'den büyükse, kelime bulunamamıştır. Diğer aday kelimenin seçilmesindeki amaç, kelime çıkartma sonucu elde edilen karakter dizisi (KD1) ile seçilen aday

kelimelerin karakter sayılarının daha yakın olanını almaktır. Böylece doğruluğu daha muhtemel kelime sınıfı seçilmiş olur. İlk karakteri atmaktaki amaç ise; bulunan karakter dizisinin (KD1) başında olmaması gereken ve daha çok ünsüzlerden oluşan karakterleri eleyerek daha doğru kelime sınıflama yapabilmektir.

Örnek vermek gerekirse; yanlış tanınan ‘alev’ kelimesi için bulunan karakter dizisi ‘haler’ olduğunda öncelikle bu kelimenin tamamı için uzaklık değeri bulunur (FD(1)). Ardından ilk karakter atılarak ‘aler’ kelimesi için uzaklık ölçüsü bulunur (FD(2)). Verilen örnekte de görüleceği gibi ‘alev’ ile ‘aler’ karakter dizileri arasındaki uzunluk ölçüsü 1’dir. Eğer sözlükte bu uzaklıktan daha küçük (0) özellik uzaklığı verecek kelime yoksa ‘alev’ kelimesine atama yapılarak işlem sonlandırılacaktır. Buna karşın sözlükte ‘alet’ kelimesi olduğu varsayıldığında uzaklık ölçüsü yine 1 bulunacak ve ‘alev’ kelimesinin tanınma olasılığı %50’ye düşecektir. Buradan da anlaşılacağı gibi kullanılan sözlüğün kelime sayısı ne kadar çoksa, birbirine yakın karakter dizileri içeren kelime sayısı artacak ve buda kelime sınıflama başarımını etkileyecektir.

BÖLÜM 6

SONUÇLAR VE ÖNERİLER

Bu çalışmada Ortak Vektör Yaklaşımı ile fonem sınıflandırılmasına dayalı, yalıtık kelime tanıma üzerinde çalışılmıştır. Çalışmada OVY'nin sürekli konuşma tanıma uygulamalarındaki kullanımını hedeflenmiştir. Bunu gerçekleştirebilmek için öncelikle yalıtık kelime tanıma üzerinde çalışılmıştır. Bu çalışmanın sonucu olarak HMM ile karşılaştırılabilir bir başarımla, ileride OVY ile oluşturulmuş, HMM'ye göre daha kullanışlı ticari bir ses tanıma uygulamasına dönüştürülebilir. Çünkü OVY'nin HMM'ye göre algoritmasının basitliği ve işlem yükü olarak üstünlükleri bulunmaktadır. Çalışmada kişi bağımlı ve kişi bağımsız kelime ve fonem tanıma sonuçları elde edilmiştir. Çalışmada sınırlı sayıda kelime içeren ses tanıma uygulamaları göz önünde bulundurularak veri tabanından kişi bağımsız kelime tanıma için 90, kişi bağımlı için 40 kelime rastgele seçilmiştir. OVY ile fonem tabanlı yalıtık kelime tanıma sonuçları, Yöntem 2A için kişi bağımsız kelime tanıma da %71.11 olarak bulunurken, fonem tanıma oranı %52.92, kişi bağımlı yalıtık kelime tanıma oranı %80 ve fonem tanıma oranı ise %64.84 olarak bulunmuştur. Yöntem 2A ve 2B ile; kişi bağımsız kelime tanıma oranı %72.22 bulunurken, fonem tanıma oranı %55.09 bulunmuştur. Kişi bağımlı yalıtık kelime tanıma oranı %82.5, fonem tanıma oranı ise %66.25'dir.

Çalışmanın çıktıları özellikle etkileşimli sesli cevap (IVR), sesle arama gibi sınırlı sayıda kelime içeren sistemlerde kullanıldığında, kelimelere ait fonem dizileri birbirine çok benzemeyeceği için yüksek başarımlar elde edilebilir. Tez kapsamında kelime çözümleme için RHA kullanılmıştır. RHA; kelimelere karşılık gelen fonem dizilerinden elde edilen karakter dizilerini kullanarak kelime eşleştirme yapan bir yöntemdir. Bu yüzden fonem sınıflama başarısı ne kadar yüksekse o kadar iyi kelime sınıflama yapmaktadır. Bunun yanında RHA'ya alternatif olarak, HMM'nin kullandığı gibi dil yapısındaki fonemlerin çıkma olasılıkları ile fonemlerin birbirini takip etme olasılıkları incelenerek oluşturulacak dil modeliyle yüksek kelime tanıma başarımları elde edilebilir. Literatür incelendiğinde HMM ile daha çok üçlü ses (triphone)

kullanılarak kelime tanıma gerçekleştirildiği görülmektedir. Türkçe’de en sık kullanılan üçlü seslerin sayısının binlerle ifade edildiği düşünüldüğünde bu üçlü seslerin OVY ile sınıflandırılması mümkün görünmemektedir. Bu yüzden çok daha az sayıdaki (38 adet) fonem sınıfının kullanımının daha isabetli olacağı görülmektedir. Ayrıca çalışmada sadece klasik OVY yöntemi kullanılmış ve altuzaylar üzerinde başarıyı artıracak çalışmalar yapılmamıştır. Tanıma başarısını artırmak için OVY metodu ile birlikte farklı altuzay teknikleri birleştirilerek ya da altuzay döndürme gibi yöntemler kullanılarak fonem sınıflama başarımları artırılabilir.

EK-A Yöntem 2A ile kişi bağımlı kelime tanıma

Kelimeler	Bulunan karakter dizileri	RHA ile tanınan kelimeler
sUryani (Süryani)	sUranıy	sUryani
kanun	karunI	kanun
bizi	vIzi	bizi
niSan (nişan)	IniSa	niSan
OrgUtU (örgütü)	vOgUtUl	OrgUtU
Ozel (özel)	vOzel	Ozel
rUSvetinin (rüşvetinin)	vrUSsetinin	rUSvetinin
Cikolata (çikolata)	viColata	Cikolata
gUller (güller)	gUler	gUller
Sekil (şekil)	Sukil	Sekil
olabilir	olavIler	olabilir
felakette	felakifer	felakette
gÜvende (güvende)	gÜveme	gÜvende
fakat	fıkak	fakat
ve	ve	ve
Uzerindeki	ozirimekil	Uzerindeki
zeki	ziki	zeki
beni	denir	beni
kim	kiSi	kim
durumu	dulumul	durumu
sanal	sala	sanal
olarak	orak	uyar
lideri	ligeyet	lideri
ziyaret	ziyeret	ziyaret
dolayI	volalil	dolayI
rahatsız	ohasuhu	kanun
tavus	kavof	kanun

Cocuk	tIyUr	CayIr
tarafIndan	tarIfimInan	tarafIndan
nazikCe	vazifUr	nazikCe
sinemasal	siyarOsal	sinemasal
uyar	vIvelar	gUvende
tavsiyeye	dasiyey	dolayI
ihiyaC	kikeliS	lideri
yallhUyUk	golulUyUy	yallhUyUk
civarInda	civana	niSan
buhar	plhar	buhar
basIncl	gasIni	basIncl
saman	salanI	saman
CayIr	Ceyer	CayIr
Tanım oranları	Fonem tanıma oranı 160/243 = %65.84	Kelime tanıma oranı 32/40 = %80

EK-B Yöntem 2A ile kişi bağımsız kelime tanıma

Kelimeler	Bulunan karakter dizileri	RHA ile tanınan kelimeler
ses	sUs	ses
hava	kova	hava
deniz	diyiv	deniz
demet	demes	demet
telefon	kilefo	telefon
iki	Uki	iki
kUCUk	kUfUk	kUCUk
Cikolata	kikOnata	Cikolata
satIcIIIk	slkIzIrak	satIcIIIk
silah	sigOr	silah
su	sU	su
kolay	kalal	alan
yUzU	dUnU	yUzU
bUtUn	gikUn	bUtUn
ve	ve	ve
Uzerindeki	ozirimekil	Uzerindeki
sIcak	sIcep	sIcak
sebep	sllep	sebep
hayIr	heyuv	hayIr
kural	kIroh	kural
biz	dUz	biz
deneme	devero	demet
sabah	salava	alan
sekizde	isikizO	sekizde
birisi	biyusa	birisi
kebap	slbap	sIcak
galip	darik	deniz

kadar	keder	kadar
kallbin	kahalun	kural
sIrada	sIlodO	sIrada
alan	anan	alan
orada	orazon	orada
Cocuk	CUcOk	Cocuk
sabit	sanuC	kanun
tipik	tIniz	deniz
olarak	olokaf	olarak
gUneS	geyiS	galip
kiSinin	kiSiy	kiSiyi
beraber	kererer	terör
kanun	kInun	kanun
da	dO	da
Ozel	gOzel	Ozel
bir	ber	bir
otoyol	ovuvol	otoyol
namaz	namOl	namaz
araba	karavav	araba
buruSuk	guSISuk	buruSuk
sonucu	soluci	sonucu
fakat	flkap	fakat
tamir	tanUy	tamir
tamamen	tanahanoC	tamamen
cesaret	cUsarul	cesaret
Celebi	kIrUvUv	kural
gUIUcUk	kUIUIUm	gUIUcUk
kar	kar	kar
Sok	Sok	Sok
cevap	cIvok	ISIk

kOtU	kUku	kanun
maCa	umoCe	maCa
her	har	her
terOr	kOyere	kOpek
badem	baderU	badem
sadece	sadize	sadece
felaket	fevevan	telefon
kiSiyi	klhUSe	kötü
yeteri	uyeter	yeteri
kadar	kOyaz	kolay
ylkama	yukano	ylkama
evet	evet	evet
keman	tenen	keten
tamam	kakar	kadar
nobel	gonan	opak
bayanın	gayalen	tamamen
mayonez	kayOyez	mayonez
kaplyı	kapi	kar
fanatik	fayatUk	fanatik
Sirin	SUren	Sirin
tarafsız	ferefas	telefon
uyarlı	eyare	uyarlı
siyah	slyaf	siyah
kİsa	kasa	kİsa
İSik	uSik	şok
hayat	kalah	sabah
kOpek	kapep	kOpek
Sehir	seyer	Sehir
kOlelik	kereyek	kOlelik
CUrUk	CUrOnu	CUrUk

keten	keyan	keman
mukus	gupOsI	uyarI
opak	kopak	opak
Tanım oranları	Fonem tanıma oranı 244/461 = %52.92	Kelime tanıma oranı 64/90 = %71,11

EK-C Yöntem 2A ve 2B ile kişi bağımsız kelime tanıma

Kelimeler	Bulunan karakter dizileri	RHA ile tanınan kelimeler
ses	tes	ses
hava	vana	hava
deniz	kiniv	İki
demet	demet	demet
telefon	tilefo	telefon
iki	lki	iki
kUCUk	kUfUk	kUCUk
Cikolata	kikOnata	Cikolata
satlcIIIk	slkIcIzak	satlcIIIk
silah	slgOr	slcak
su	sU	su
kolay	volale	kolay
yUzU	yUnU	yUzU
bUtUn	bUkUn	bUtUn
ve	ve	ve
Uzerindeki	ozirimekil	Uzerindeki
slcak	sahap	sabah
sebep	sllep	sebep
haylr	hayel	haylr
kural	flloh	silah
biz	diz	biz
deneme	develÖ	evet
sabah	solav	silah
sekizde	sekazO	sekizde
birisi	bivisa	birisi
kebab	sarap	sabah

galip	karOk	kUCUk
kadar	keder	kadar
kallbın	kanupun	kanun
sIrada	sIvodo	sIrada
alan	anan	alan
orada	orazö	orada
Cocuk	COcOk	Cocuk
sabit	sanuC	kanun
tipik	fUpizi	tipik
olarak	olovok	olarak
gUneS	gUkUh	gUneS
kiSinin	keSe	keten
beraber	kererer	terör
kanun	kInun	kanun
da	dO	da
Ozel	lOzel	Ozel
bir	ber	bir
otoyol	opöyol	otoyol
namaz	namol	namaz
araba	valolo	alan
buruSuk	kuSuSuk	buruSuk
sonucu	soluci	sonucu
fakat	fIkat	fakat
tamir	tamUg	tamir
tamamen	tamareC	tamamen
cesaret	CUsOrUt	cesaret
Celebi	sIrUvUv	sIrada
gUIUcUk	gUcUIUm	gUIUcUk
kar	kar	kar
Sok	Cok	Sok

cevap	clvok	ISIk
kOtU	kItI	kOtU
maCa	umoCe	maCa
her	her	her
terOr	kOrere	kOpek
badem	baderU	badem
sadece	sadize	sadece
felaket	fevekan	telefon
kiSiyi	heSe	her
yeteri	uyeter	yeteri
kadar	kOyaz	kolay
ylkama	yukamo	ylkama
evet	evet	evet
keman	teman	keman
tamam	vamam	tamam
nobel	gonan	kolay
bayanın	bayalen	bayanın
mayonez	kayOyez	mayonez
kapIyı	kapi	kar
fanatik	fayatUk	fanatik
Sirin	Siren	Sirin
tarafsIz	ferefas	telefon
uyarI	eyare	uyarI
siyah	siyaf	siyah
kIsa	kısar	kIsa
ISIk	aSik	Sok
hayat	hanan	hayat
kOpek	tepek	kOpek
Sehir	Seyer	Sehir
kOlelik	veyeyek	deneme

CUrUk	CerOni	Celebi
keten	vOkan	keman
mukus	mupOtI	uyarI
opak	kopak	opak
Tanım oranları	Fonem tanıma oranı 254/461 = %55.09	Kelime tanıma oranı 65/90 = %72.22

EK-D Yöntem 2A ve 2B ile kişi bağımlı kelime tanıma

Kelime	Bulunan Karakter Dizisi	RHA ile Tanınan Kelime
sUryani (Süryani)	sUranıy	sUryani
kanun	kanInI	kanun
bizi	bizi	bizi
niSan (nişan)	InıSa	niSan
OrgUtU (örgütü)	vOgUtUl	OrgUtU
Ozel (özel)	vOzel	Ozel
rUSvetinin (rüşvetinin)	lurUSetinin	rUSvetinin
Cikolata (çikolata)	CiColaka	Cikolata
gUller (güller)	gUler	gUller
Sekil (şekil)	Sekil	Sekil
olabilir	olabIler	olabilir
felakette	felakefer	felakette
gUvende (güvende)	gUveme	gUvende
fakat	farat	fakat
ve	ve	ve
Uzerindeki	ezirimikiy	Uzerindeki
zeki	ziki	zeki
beni	beniy	beni
kim	kiSi	kim
durumu	dulumul	durumu
sanal	sala	sanal
olarak	orak	uyar
lideri	ligeret	lideri
ziyaret	ziyeret	ziyaret
dolayI	doloyir	dolayI
rahatsız	ahotohu	dolayI

tavus	tavof	tavus
Cocuk	voclp	Cocuk
tarafIndan	tarIfinan	tarafIndan
nazikCe	dalUr	CayIr
sinemasal	silabasal	sinemasal
uyar	vlyelar	ve
tavsiyeye	vasiyef	nazikCe
ihitiyaC	gikeliS	lideri
yalIhUyUk	bOlulUyUk	yalIhUyUk
civarInda	cUrUvana	sUryani
buhar	bohar	buhar
basInCI	basIci	basInCI
saman	salanI	saman
CayIr	Ceyir	CayIr
Tanım oranları	Fonem tanıma oranı 161/243 = %66.25	Kelime tanıma oranı 33/40 = %82,5

EK-E 60, 80, 100,120 ve 150 ms uzunluklu 32 fonem sınıfı için bulunan başarımlar

80 ms	Eğitim (91) *		Eğitim (84)		Test (91) *		Test (84)	
	U	UZ	U	UZ	U	UZ	U	UZ
KB	100	100	100	100	39.75	49.37	39.08	48.31
KBZ	82.27	85.35	80	85.4	52	57.3	50.6	57.5

150 ms	Eğitim (182) *		Eğitim (168)		Test (182) *		Test (168)	
	U	UZ	U	UZ	U	UZ	U	UZ
KBZ	99.9	99.87	99.7	99.87	40.3	54.32	42.1	55.62

120 ms	Eğitim (143) *		Eğitim (132)		Test (143) *		Test (132)	
	U	UZ	U	UZ	U	UZ	U	UZ
KBZ	98	99.03	97.07	98.79	41.1	55.81	40.3	53.91

100 ms	Eğitim (117) *		Eğitim (108)		Test (117) *		Test (108)	
	U	UZ	U	UZ	U	UZ	U	UZ
KBZ	93.7	97	91.2	95.3	40.27	55.06	40.15	52.98

80 ms	Eğitim (91) *		Eğitim (84)		Test (91) *		Test (84)	
	U	UZ	U	UZ	U	UZ	U	UZ
KBZ	82.27	85.35	80	85.4	39.75	49.37	39.08	48.31

60 ms	Eğitim (65) *		Eğitim (60)		Test (65) *		Test (60)	
	U	UZ	U	UZ	U	UZ	U	UZ
KBZ	64.4	68.1	60.1	62.89	36.55	42.07	36.1	42.23

EK-F 60, 80, 100,120 ve 150 ms uzunluklu 27 fonem sınıfı için bulunan başarımlar

80 ms	Eğitim (91) *		Eğitim (84)		Test (91) *		Test (84)	
	U	UZ	U	UZ	U	UZ	U	UZ
KB	85.7	90	83.3	87.6	48.75	53.02	45.51	50.75
KBZ	100	100	100	100	63.3	63.06	65	61.03

150 ms	Eğitim (182) *		Eğitim (168)		Test (182) *		Test (168)	
	U	UZ	U	UZ	U	UZ	U	UZ
KBZ	99.96	99.95	99.95	99.93	49	58.85	48.46	57.26

120 ms	Eğitim (143) *		Eğitim (132)		Test (143) *		Test (132)	
	U	UZ	U	UZ	U	UZ	U	UZ
KBZ	98	99.3	97.3	99.1	48.1	58.2	46.6	56.77

100 ms	Eğitim (117) *		Eğitim (108)		Test (117) *		Test (108)	
	U	UZ	U	UZ	U	UZ	U	UZ
KBZ	94.84	97.5	93.1	96	47.35	56.97	45.95	55.8

80 ms	Eğitim (91) *		Eğitim (84)		Test (91) *		Test (84)	
	U	UZ	U	UZ	U	UZ	U	UZ
KBZ	85.7	90	83.3	87.6	48.75	53.02	45.51	50.75

60 ms	Eğitim (65) *		Eğitim (60)		Test (65) *		Test (60)	
	U	UZ	U	UZ	U	UZ	U	UZ
KBZ	70	70.3	66.6	65.5	45.72	46.67	45.05	45.82

KAYNAKLAR DİZİNİ

- Abdulla, W.H., Chow, D. and Sin, G. Cross-words Reference Template for DTW-based speech recognition systems, Conference on Convergent Technologies for Asia-Pacific Region, p: 1576 - 1579 ,2003.
- Arısoy, E. ve Arslan, L.M.,2005,Turkish Dictation Sytem for Broadcast News Applications,Signal Processing and Communications Applications Conference, Proceedings of the IEEE 13th Volume , Issue , p: 629 – 632
- Arısoy, E., Dutagaci, H. and Arslan, L. M., 2006, A unified language model for large vocabulary continuous speech recognition of Turkish, Signal Processing, p: 2844-2862.
- Arslan, L.M., 1999 ,Türkçe Sürekli Konuşma Tanıma Sisteminin Sayı Tanıma Uygulaması, Sinyal İşleme ve İletişim Uygulamaları Kurultayı,Ankara, s: 64-67.
- Bayrakçeken, M.K.,2005, Ortak Vektör Yaklaşımı Kullanılarak Kesintisiz Konuşma Metinlerinde Anahtar Kelime Bulma”,Yüksek lisans tezi, Eskişehir, Osmangazi Üniversitesi.
- Çevikalp, H. , Neamtu, M. ,Wilkes, M. and A. Barkana, 2005,Discriminative common vectors for face recognition,IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no. 1, p. 4-13.
- Çölkesen, R., 2004 ,Veri Yapıları ve Algoritmalar, Papatya Yayıncılık, 424 s.
- Demirci,M.D. ve Öztaş O. 2005,MFCC ve Vektör Kuantizasyonu Kullanılarak Konuşmacı Tanıma, İstanbul Üniversitesi, Bilgisayar Mühendisliği Bölümü, MEGAK.
- Edizkan, R., Tiryaki, B., Büyükcan, T. ve İsmail UZUN, 2007,Ses Komut Tanıma İle Gezgin Araç Kontrolü’ Eskişehir Osmangazi Üniversitesi,Elektrik Elektronik Mühendisliği Bölümü.
- Erickson, J., Hash Tables, University of Illinois,Lecture 6, 2002.
- Furui, S., 1989. Digital Speech Processing, Synthesis, and Recognition. Marcel Dekker Inc.

KAYNAKLAR (Devam)

- Furui, S.; Kikuchi, T., Shinnaka, Y and Hori, 2004, C ,Speech-to-text and speech-to-speech summarization of spontaneous Speech, Speech and Audio Processing, IEEE Trans. On Volume 12, Issue 4, p: 401 – 408.
- Gorin, A. and Mammone, R.J. , 1994 Introduction to the Special Issue on Neural Networks for Speech Processing, Speech and Audio Processing, IEEE Transactions on Volume 2, Issue 1, p:113 – 114.
- Günel, S. and Edizkan, R. 2007,Use of Novel Feature Extraction Technique with Subspace Classifiers for Speech Recognition, EEE International Conference on Volume , Issue , 15-20 , p:80 – 83.
- Gülmezoğlu, M. B. , Dzhafarov, V. , Edizkan R. and Barkana, A. 2007, The Common Vector Approach and Its Comparison with Other Subspace Methods in Case of Sufficient Data,Computer Speech and Language, Vol.(21) 2, p: 266-281.
- Gulmezoglu, M.B, Dzhafarov, V.,Keskin, M. and Barkana, A. , 1999 A Novel Approach to Isolated Word Recognition, Speech and Audio Processing, IEEE Transactionson Volume 7, Issue 6, p: 620 – 628.
- Gulmezoglu, M.B., Dzhafarov, V.and Barkana, 2001, A. The Common Vector Approach and its Relation to Principal Component Analysis' Speech and Audio Processing, IEEE Transactions on Volume 9, Issue 6, p:655 – 662.
- Karjanadecha M. and Zahorian, S.,2001,Signal Modeling for High-Performance Isolated Word Recognition, IEEE Transactions on Speech and Audio Processing , p: 647-654.
- Kohonen, T., Riittinen, H., Jalanko, M., Reuhkala, E. and Haltsonen, S., 1980., A thousand-word recognition system based on the learning subspace method and redundant hash addressing. In: Proceedings of the Fifth International Conference on Pattern Recognition, p: 158-165.
- Kurcan R.S., M.Fargues and R. Vaidyanathan 2006,In-Ear Microphone Speech Data Recognition using HMMs, Digital Signal Processing Workshop, 12th - Signal Processing Education Workshop, 4th Volume p: 268 – 272.
- Mantysalo, J., Tarkkolat, K. and Kohonen, T., 1992.,LVQ-based Speech Recognition with High Dimensional Context Vectors, Helsinki University of Technology, Laboratory of Information and Computer Science. p: 539-542.
- Mengüşoğlu, E.1999, Bir Türkçe Sesli İfade Tanıma Sisteminin Kural Tabanlı Tasarımı ve Gerçekleştirimi, Hacettepe Üniversitesi, Yüksek Mühendislik Tezi.

KAYNAKLAR (Devam)

- Rabiner, L., R and Schafer, R., W., 1978. Digital processing of speech signals, Prentice Hall Inc.
- Reuhkal, E., Jalanko, M. and Kohonen, T ,1979 ‘A Radundant Hash Addressing Method Adapted For The Postprocessing and Error-Correction of Computer-Recognized Speech ,Helsinki University of Technology p: 591-594.
- Salor, Ö., Pellom, B., Çiloğlu, T., Hacıoğlu, K. and Demirekler, M.,2002, New Corpora and Tools for Turkish Speech Research, Sinyal İşleme ve İletişim Uygulamaları Kurultayı, s: 476-481.
- Selen, N., 1979, Söyleyiş sesbilimi, akustik sesbilim ve Türkiye türkçesi, Türk Dil Kurumları Yayınları.
- SESTEK IVR (Etkileşimli Sesli Yanıt Sistemi). 2008_ SesTek Ses ve İletişim Bilgisayar Teknolojileri San. ve Tic. A.Ş. , <http://www.sestek.com.tr> .
- Somervuo ,P. , 2004, Redundant Hash Addressing of Feature Sequences using the Self-Organizing Map’, Helsinki University of Technology, Finland.
- Turk O.ve Arslan L.M.,2004., Konuşma Terapisine Yönelik Otomatik Konuşma Tanıma Yöntemleri, II. Ulusal Dil ve Konuşma Bozuklukları Kongresi, Eskişehir.
- Yağmur, O., (11 Mart 2007) ,Hashing Fonksiyonları, Hash Tablosu ve Collision. <http://www.csharpnedir.com>.