

T.C.

771109

GEBZE YÜKSEK TEKNOLOJİ ENSTİTÜSÜ
MÜHENDİSLİK VE FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKÇE OTOMATİK DUDAK
OKUMA SİSTEMİ

Alper COŞKUN
YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ
ANABİLİM DALI

TEZ DANIŞMANI
Yrd. Doç. Dr. Mehmet GÖKTÜRK

GEBZE
2005



YÜKSEK LİSANS JÜRİ ONAY FORMU

GEBZE YÜKSEK TEKNOLOJİ
ENSTİTÜSÜ

Alper CoşKUN'un tez çalışması, G.Y.T.E. Mühendislik ve Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 219/08/2005 Tarih ve 2005/27 sayılı kararıyla oluşturulan jüri tarafından Bilgisayar Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : Yrd.Doç.Dr. Mehmet GÖKTÜRK

ÜYE

: Yrd.Doç.Dr. Yusuf Sinan AKGÜL

ÜYE

: Yrd.Doç.Dr.İlyas KANDEMİR

ONAY

G.Y.T.E. Mühendislik ve Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
24.11.2005 Tarih ve 2005/40 sayılı kararı.

İMZA/MÜHÜR

ÖZET

Teknolojideki ilerleme, otomatik konuşmayı anlama sistemlerinin Bilgisayar-İnsan Etkileşimi'nde yaygın bir şekilde kullanılmaya başlanmasına neden oldu. Bununla birlikte bu sistemler hala sıradan ortamlarda yetersiz kalmaktadır. Şu anki uygulamalar; ofis, açık hava gibi ortamlarda rastlanılan arkaplan gürültüsü veya birden fazla insanın aynı anda konuştuğu durumlara karşı çok hassastırlar. Bu durumun üstesinden gelmenin yollarından biri, konuşmayı anlama sistemlerinde insanların yaptığı gibi görsel bilgilerden yararlanmaktır. Araştırmalara göre insanların konuşmayı anlamasında görsel işaretler önemli rol oynamaktadır. Bu yüzden otomatik konuşmayı anlama sistemlerinde dudak okumadan da yararlanılması, bizi daha verimli Bilgisayar-İnsan Etkileşimi'ne bir adım daha yaklaştıracaktır. Bu projenin temel hedefi, Türkçe diline dayalı uygulanabilir bir dudak okuma sistemi oluşturmaktır. Türkçe dili için iki işitme engelli tarafından bir fonem-vaysem eşleşim tablosu hazırlanmış ve dudak biçimini çıkaran bir sistem oluşturularak, bu eşleşim tablosuyla birlikte, konuşmacının ne dediğini anlamaya çalışan bir uygulama oluşturulmaya çalışılmıştır.

SUMMARY

Advances in technology in recent years have led to a widespread use of automatic speech recognition (ASR) systems in Human-Computer Interaction (HCI). However, ASR systems still perform poorly, when it comes to real world applications. Most approaches are very sensitive to background noise or fail when more than one speaker talks simultaneously, as it often happens in offices, outdoors and other real world environments. One way of overcoming these limitations of audio-only ASR systems is to use the additional visual information of the act of speaking, just like humans do. Psycholinguistic research has found that visual cues like lip movements play an important role in speech understanding by humans. Thus, a lip reading system as part of an ASR system leads thus one step closer to more natural HCIs. The major aim of this project is to show a simple and applicable way of a "Lip Reading System" on Turkish Language. Two hard-of-hearing people have prepared a Turkish language phoneme-viseme mapping table. Some basic lip features are extracted. After that; these features and the mapping table are used as to find that what the speaker says.

TEŞEKKÜR

Bu çalışmaya olan katkıları ve değerli önerileri nedeniyle danışmanım Yrd. Doç. Dr. Mehmet Göktürk'e teşekkür ederim.



İÇİNDEKİLER DİZİNİ

	<u>Sayfa</u>
ÖZET	iv
SUMMARY	v
TEŞEKKÜR	vi
İÇİNDEKİLER DİZİNİ	vii
ŞEKİLLER DİZİNİ	viii
ÇİZELGELER DİZİNİ	ix
1. GİRİŞ	1
2. DUDAK OKUMAYLA İLGİLİ GENEL BİLGİLER	3
2.1. Vaysem ve Fonem	3
3. TÜRKÇE EŞLEŞİM TABLOSU	5
4. GÖRÜNTÜDEN DUDAK BİÇİMİNİN ÇIKARIMI	8
4.1. Yatay Sınırlara Dayalı Çıkarım	8
4.2. KYM'ye Dayalı Çıkarım	8
4.3. RDY'ye Dayalı Çıkarım	9
4.4. Kırmızının İhmaline Dayalı Çıkarım	10
5. DUDAK BİÇİMİNİN SAYISAL İFADESİ	12
6. DEĞERLENDİRME	15
7. İMPLEMENTASYON DETAYLARI	17
8. SONUÇ	18
KAYNAKLAR	20
ÖZGEÇMİŞ	22

ŞEKİLLER DİZİNİ

<u>Sekil</u>	<u>Sayfa</u>
3.1. Dudak Okuma Aşamaları.	7
4.1. Programdan Örnek Bir Ekran Görüntüsü.	11
5.1. İmge Bölgesi Şablonu.	12
5.2. Dudak Biçiminin Sayısal İfadesi	14
6.1. Verilerin Harflere Dönüşümü	15



ÇİZELGELER DİZİNİ

Cizelge

Sayfa

3.1 Türkçe Vaysem-Fonem Eşleşim Tablosu.

6

5.1. İmge Bölgesi Algoritmasının Basamakları.

13

8.1 Karşılaştırma

19



1. GİRİŞ

Bilgisayar-İnsan etkileşiminde temel araçlar olan klavye, fare ve benzerlerine yeni etkileşim araçları katmak için çalışmalar uzun yıllardır devam etmektedir. Daha etkin ve hızlı etkileşim yollarına ulaşmaya çalışılmaktadır. Bu yönde yapılan çok çeşitli çalışmaların başında “konuşmayı tanıma” gelmektedir. İnsan konuşmasını metne dökabilen (speech recognition) ve insan konuşmasını taklit edebilen (text-to-speech) bu sistemler üzerinde yapılan yoğun çalışmalar sonucunda, yeterli verim alınabilen yazılımlar üretilmiştir. Örneğin Microsoft şirketi, en son işletim sistemi Windows XP’ye “Microsoft Speech” adını verdiği bir konuşmayı tanıma sistemi eklemiştir [Microsoft, 2005]. Diğer bir örnek ise, güncel cep telefonlarının çoğunda ses ile arama desteği olmasıdır. Bunların yanı sıra çeşitli nedenlerle el, göz ve dudak gibi diğer organların da takibini ve yorumunu yapmaya dayalı çalışmalar yapılmaktadır.

Öte yandan, bu konuşmayı tanıma sistemleri ancak ideal ortamlarda istenilen başarıya ulaşmaktadır. Günlük yaşamda sıklıkla bulunduğumuz gürültülü ortamlarda veya birden fazla kişinin aynı anda konuştuğu durumlarda başarı yüzdeleri düşmektedir [Bregler et. al., 1993]. Bu durumla başa çıkmanın yollarından biri, işitme engellilerin de faydalandığı dudak okuma yöntemini kullanmaktır.

İşitme engellilerin, dudak okuyarak konuşmaları anlayabildiği bilinen bir gerçektir. Dudak okuma; sestten yararlanmadan veya az yararlanarak, dudak hareketlerinin takibiyle insanların konuşmasını anlamaya dayalı bir yöntemdir. İşitmesi normal insanlar da, çoğu zaman farkında olmadan dudaktan okuma yönteminden yararlanmaktadırlar [Dodd and Campbell, 1987]. Fakat bu işi en iyi başarabilenler sağırılar değil, işitme yetisini tamamen kaybetmemiş olan işitme engellilerdir. Çünkü dudaktan okuma ses ile desteklenince daha verimli sonuçların alındığı bilinmektedir. Araştırmalara göre sesin hiç iletilmediği bir durumda, dudaktan okuma ile söylenenleri anlama yeteneği ortalama %40-50’yi geçmemektedir [UCSC, 1996; Wikipedia, 2005].

Dudaktan okumada tahmine sıklıkla başvurulur. Çünkü söylenen sözlere karşılık gelen dudak hareketleri kişiden kişiye farklılıklar gösterebilir. Ayrıca harflere değil, kelimelere hatta kelime gruplarına dayalı bir yöntemdir [UCSC, 1996]. Harflerin tek tek takibinin güç olacağı açıktır. Bunların yanı sıra bir cümle içinde kaçırılan bazı kelimeler için sıklıkla tahmine başvurulur. İşitme engelliler, bu yüzden konuşmayı takip ederken, tüm söylenenleri tek tek anlamaktan ziyade, konuşmanın içeriğini anlamaya odaklanırlar. Bu durumda, konuşmanın ana konusunun ne olduğunu önceden bilmek onlar için faydalıdır.



2. DUDAK OKUMAYLA İLGİLİ GENEL BİLGİLER

İnsan konuşmasının anlamlı hale gelmesi için, sadece ses yeterli değildir. Konuşmanın biçimlenmesinde gırtlak, dişler, dil ve dudaklar da rol oynar. Bunlardan dışarıdan en fazla görüneni dudaklardır. Dişler ve dil de çoğu zaman görünmesine rağmen, konuşmayla ilgili bize dudaklar kadar bilgi vermezler. Bu nedenle bu yönteme dudak okuma denmektedir.

Dudak okumanın bilgisayar ortamında yapılabilmesi için dudakların iyi ışıklandırılmış bir ortamda kamera ile gerçek zamanda net bir şekilde kaydedilmesi; bu kayıttaki görüntüyü oluşturan resimlerin içinde dudakların o anki durumlarının sayısal olarak algoritmalarla saptanması gerekir. Daha sonra bu veriler yorumlanarak, kelimelere dönüştürülmeye çalışılır.

Dudak okuma konusunda ilk yapılan çalışma Petajan adlı doktora öğrencisi tarafından 1984 yılında yapılmıştır. Bu çalışmada Petajan, sadece beyaz ve siyah renklerden oluşan bir görüntüden dudak biçimini çıkarmayı başarmış, sadece görsel verileri işleyen ve “dinamik zaman sarınımına” (dynamic time warping) dayanan bir tanıma sistemi geliştirmiştir [Petajan, 1984].

Dudak okumayla ilgili araştırmalar genelde İngilizce, Almanca, Fransızca gibi diller için yapılmıştır. Ulaşabildiğimiz kaynaklara göre, şimdiye kadar yapılmış Türkçe bir çalışma yoktur.

2.1. Vaysem ve Fonem

Fonem (phoneme), özgün bir sestir. Her dilde kendine özgü fonemler vardır. Fonemler birleşerek anlamlı kelimeler oluştururlar. Bir kelimeyi oluşturan fonemlerin birinin bile değişmesi, kelimeyi tamamen değiştirebilir. Fonemin görsel karşılığı ise vaysem (viseme) olarak adlandırılır. Dudak okumada bu vaysemlemler ile

fonemlerin eşleřtirildięi tablolar kullanılır [Kleeman, 1994]. Tabloda belli bir fonem, örneęin “a” ve ona karřılık gelen dudak biçimi detaylı bir řekilde yazılır.

Genellikle vaysem için; dudakların açık olup olmadıęı, dudakların řekli (yuvarlak, düz, üflermiř gibi), diřlerin durumuna göre bir tarif yapılır. Ayrıca bu bilgiler dudaęın geniřlięi, yükseklięi, çevresi ve alanı olarak sayısal açıdan da ifade edilebilirler.



3. TÜRKÇE EŞLEŞİM TABLOSU

Ana dili Türkçe olan iki işitme engellinin ortak çalışmasıyla, örnek bir Türkçe vaysem-fonem eşleşim tablosu oluşturulmuştur. Her iki işitme engellinin de dudak okuma yetenekleri oldukça gelişmiştir ve engelleri sonradan oluştuğu için sesleri de tanımaktadırlar. Bu yüzden bu çalışma için ideal adaylardır. Çalışmayı yapmadan önce, bu konuda başka bir çalışma olup olmadığı araştırılmıştır. Ulaşabildiğimiz kaynaklara göre, Türkçe için böyle bir çalışma şimdiye kadar yapılmamıştır. Bununla birlikte, dilin özelliklerinden dolayı, birkaç fonemin hangi vaysemlere karşılık geldiği çeşitli Türkçe dilbilgisi kitaplarında ve konuşmayı tanıma ile ilgili bir çalışmada [Türk ve Arslan, 2004] yer almaktadır. Fakat bu birkaç fonemin bu çalışmayı gerçekleştirmek için yetersiz kalacağı açıktır. Ayrıca yabancı dillerde benzer çalışmaların nasıl yapıldığı incelenmiştir [Visser, 1999].

Bu konuda yararlanılan diğer bir kaynak ise SAMPA (Speech Assessment Methods Phonetic Alphabet) 'tir [SAMPA, 2002]. Bu listede Türkçe dahil olmak üzere bir çok yaygın dilin “bilgisayar tarafından yorumlanabilir” (computer-readable) şekillerine yer verilmiştir. Listede fonemler, duruş sessizi (plosive), zorlayıcı sessiz (fricative), geniz sessizi (nasal) ve akıcı sessiz (liquid) gibi gruplara ayrılmıştır.

İki işitme engelli de belirtilen kaynaklardan yararlanıp, ayrı ayrı çalışarak bir eşleşim tablosu oluşturduktan sonra; bu iki çalışma birleştirilmiştir. Böylelikle, daha uygulanabilir ve güvenilir bir sonuca ulaşma şansının arttırıldığı düşünülmektedir.

Türkçe fonetik yani yazıldığı gibi okunan bir dil olduğu için, fonemlerinin sayısı fazladır. Bununla birlikte bazı fonemler birbirine çok yakındır. Vaysemler açısından ayırt edici özellikleri en fazla olan fonem-vaysem eşleşmeleri saptanmaya çalışılmıştır. Çizelgede görülebileceği gibi sonuç olarak 15 foneme karşılık gelen, 9 ayrı vaysem saptanmıştır.

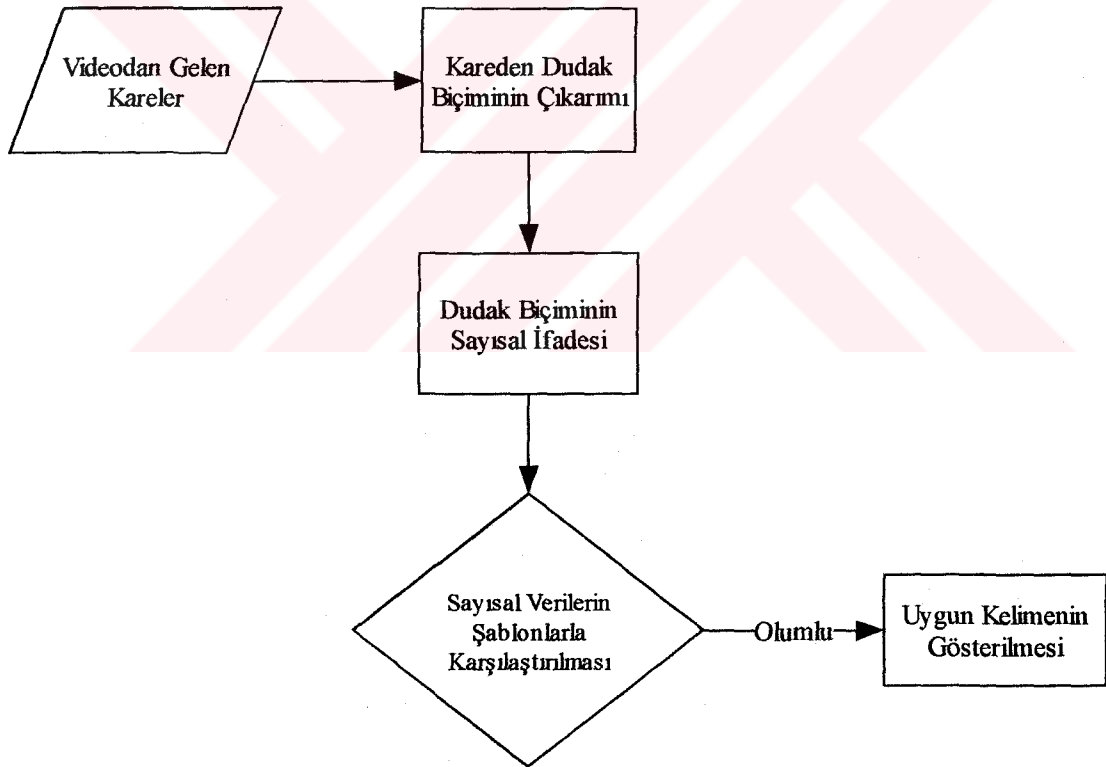
Çizelge 3.1 Türkçe Vaysem-Fonem Eşleşim Tablosu.

Türkçe Vaysem-Fonem Eşleşim Tablosu		
Harf	Tanım	Detaylı Tanım
a	Sesli	Dudaklar açık, alt dudak aşağıya doğru
e	Sesli	Dudaklar açık, yanlara doğru gerilmiş
ı-i	Sesli	Dudaklar az açık, dişler görünüyor
o-ö	Sesli	Dudaklar “o” şeklinde
u-ü	Sesli	Dudaklar üfürürmüş gibi
l	Dil Sessizi	Dudaklar normal miktarda açık
b-m	Dudak Sessizi	Dudaklar kapalı
f-v	Diş-Dudak Sessizi	Diş ve dudak birbirine değiyor
p	Çift Dudak Sessizi	Dudaklar kapalı

DUDAK OKUMA AŞAMALARI

Bilgisayar ile dudak okuma, *dudakların o anki biçimlerinin sayısal olarak ifade edilmesi* ve *sayısal verilerin kelimeler olarak yorumlanması* olmak üzere iki temel aşamadan oluşur [Visser, 1999].

Biçimlerin sayısal olarak ifade edilmesinde görüntünün kameradan kare kare alınması, alınan karelerde dudakların çıkarımı, ve dudakların biçimlerinin sayısal veriler olarak elde edilmesi olmak üzere iki alt aşama vardır. Bu sayısal verilerin kelimeler olarak işlenmesi ise; elde edilen sayısal dudak verilerinin önceden hazırlanmış tabloyla karşılaştırılması ve uyumlu olduğu saptanan kelimelerin ekranda gösterilmesi olmak üzere iki alt aşamadan oluşur. Aşağıdaki şekilde tüm bu aşamalar topluca gösterilmeye çalışılmıştır.



Şekil 3.1. Dudak Okuma Aşamaları.



4. GÖRÜNTÜDEN DUDAK BİÇİMİNİN ÇIKARIMI

Görüntüden dudağın çıkarımıyla ilgili farklı yöntemler mevcuttur. Bunlar temel olarak dört grupta toplanabilir:

- Yatay Sınırlara Dayalı
- KYM (Kırmızı, Yeşil, Mavi)'ye Dayalı
- RDY (Renk Özü, Doygunluk, Yoğunluk)'a Dayalı
- Kırmızının İhmaline Dayalı

Bu yöntemlerin hepsi düşük-seviye olarak adlandırılır. Dudağın geometrik özelliklerini tespit etmeye yöneliktirler. Yüksek-seviye dudak çıkarımı yöntemleri ise, dudağın belli şekillerde modellenmesine dayanır. Yüksek-seviye yöntemlerine örnek olarak *snakes* [Chiou et. al., 1997] verilebilir.

4.1. Yatay Sınırlara Dayalı Çıkarım

Bu yöntem gri renk dağılımına ve sınır saptamaya (edge detection) dayalıdır [Steifelhagen et. al., 1997]. Bu yöntemin temelinde ağız bölgesinin keskin yatay çizgilere sahip olduğu düşüncesi yatar. Genelde iyi sonuç verebilen bu yöntem, sakal veya bıyık olduğu durumlarda başarısız sonuç verir [Lewis and Powers, 2001].

4.2. KYM'ye Dayalı Çıkarım

Araştırmalarda yüz bölgesinin konumunun saptanması gerekliliğini kaldıran bu yöntemde dudak bölgesini ve özelliklerini bulmak için, renk spektrumuna bakılır. Örnek olarak Wark ve ark., şu denklemi kullanmışlardır:

$$A_{sum} \leq \frac{K}{Y} \leq \ddot{U}_{sum}$$

Burada K, kırmızı, Y, yeşil renk değerleridir. A ve $\ddot{U}$ ise dudakların bulunduğu aralığı belirler [Wark et. al., 1998].

Alakasız piksellerin (gürültü) filtrelenmesinden ve morfolojik açma-kapama işlemlerinden sonra yukarıdaki denklem uygulandığında genellikle “yatay sınırlara dayalı çıkarım”dan daha iyi sonuç vermektedir. Fakat bazı durumlarda dudakların köşelerini, bazı durumlarda ise dudakların üst ve alt kenarlarının saptanmasında güçlük olmaktadır [Lewis and Powers, 2001].

4.3. RDY’ye Dayalı Çıkarım

Renk özü (hue), doygunluk (saturation) ve yoğunluk (intensity) değerlerine dayalı çıkarımın ana fikri; renk özü değerinin ortam koşullarından etkilenmemesidir. Örneğin ortam ışığının az miktarda artması veya azalması, renk özü değerine çok az etki eder. Belli bir pikselin dudağın bir parçası olma ihtimalini hesaplayan aşağıdaki denklem Coianiz ve ark. [Coianiz et. al., 1996] tarafından kullanılmıştır:

$$f(h) = \begin{cases} 1 - \frac{(h - h_0)^2}{w^2}, & |h - h_0| \leq w \\ 0, & otherwise \end{cases}$$

h_0 önceden hesaplanmış bir renk özü değeri, h o anki renk özü değeri, w ise çevredeki değeri sıfır ve sıfıra yakın piksellerin uzaklığıdır.

Bu yöntemin dezavantajı ağız bölgesi dışındaki tene ait kısımların da bazı durumlarda yanlış bir şekilde ağızın bir parçasıymış gibi saptanmasıdır [Lewis and Powers, 2001].

4.4. Kırmızının İhmaline Dayalı Çıkarım

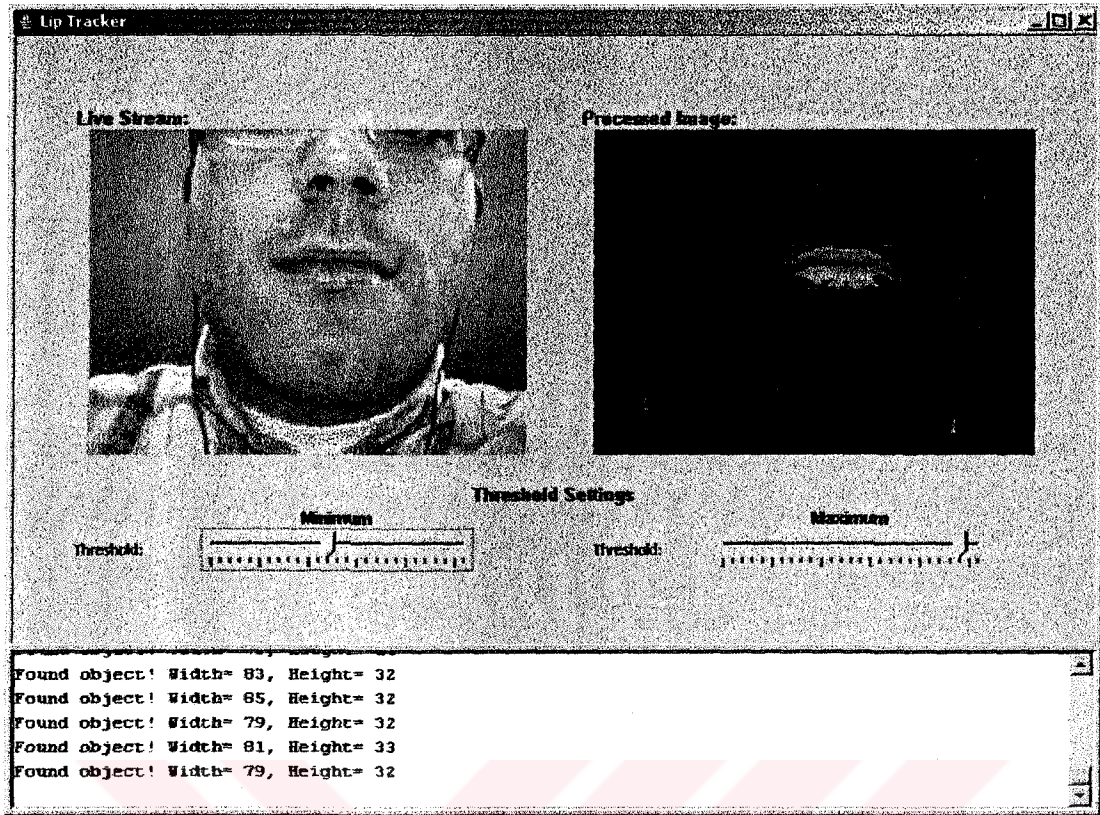
Bu yöntem kırmızıya bakmak yerine, mavi ve yeşil değerlere bakar. Ana fikir; yüzün ve dudakların, baskın bir şekilde kırmızı renk ve tonlarından oluşmasıdır. Bu yüzden bu bölgeler arasındaki asıl farklar yeşil ve mavi renk aralıklarında yer alır [Lewis and Powers, 2001]. Dudaklar ile yüz bölgesi arasındaki renk farkı ise; yüz bölgesinde yeşil rengin daha fazla olması, dudak bölgesinde ise yeşil ve mavi rengin birbirine yakın olmasıdır.

Resimde yer alabilecek alakasız pikseller (gürültü) filtrelendikten sonra, resim aşağıdaki denkleme dayalı olarak eşik işleminden (thresholding) geçirilir.

$$-\beta \leq \log\left(\frac{G}{B}\right) \leq \beta$$

Eşik işleminden sonra, resim morfolojik açma ve kapama işleminden geçirilir. Böylece ufak boşluklar ve ufak lekeler resimden kaldırılmış olur. Son olarak resim sadece beyaz ve siyah renkleri içerecek biçime (binary / ikili resim) dönüştürülür.

Bu çalışmada “Kırmızı’nın İhmaline” ve “RDY’ye Dayalı” yöntemler beraberce kullanılmıştır. Şekil 4’te programdan alınmış bir ekran görüntüsüne yer verilmiştir.

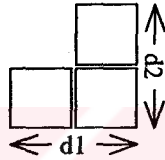


Şekil 4.1. Programdan Örnek Bir Ekran Görüntüsü.

5. DUDAK BİÇİMİNİN SAYISAL İFADESİ

Dudak biçimi belirlendikten sonra bulunduğu konumun sayısal olarak saptanması gerekir. Bunun için “imge bölgesi” (blob) yöntemi kullanılmıştır [Perez, 2003]. Bir imge bölgesi, birbirine bir komşuluk ilişkisiyle bağlı bir grup piksele verilen addır [Gonzalez and Woods, 2001]. Bu çalışmada dudakların komşuluk ilişkisi ise piksellerin aynı renkte olmalarıdır.

Söz konusu algoritma, resim üzerinden şekilde görüldüğü gibi ters L biçimindeki bir şablonun geçirilmesine dayanır.



Şekil 5.1. İmge Bölgesi Şablonu.

İmge Bölgesi Şablonu, o anki piksel ile solundaki ve üstündeki pikseller arasındaki mesafeyi hesaplamak için kullanılır. KYM (Kırmızı, Yeşil, Mavi) renk paletine dayalı bir resimde bu mesafe aşağıda görülen öklid mesafesi formülü ile hesaplanır [Bull, 2005].

$$KYM_Öklid_Mesafesi = \sqrt{(r_1 - r_2)^2 + (g_1 - g_2)^2 + (b_1 - b_2)^2}$$

Resmin KYM renk paletinin RDY (Renk özü, Doygunluk, Yoğunluk) paletine geçirilmesi halinde ise, sadece “renk özü” değeri mesafesinin saptanması yeterlidir.

Önceden saptanmış eşik değeri (P) ile üst ve sol piksellerin karşılaştırılması Çizelge 2’de gösterildiği gibi yapılır.

Çizelge 5.1. İmge Bölgesi Algoritmasının Basamakları.

Durum	İzlenecek Yol
$(d_1 > P) \text{ ve } (d_2 > P)$	O anki piksel iki komşusundan da farklı, demek ki yeni bir bölgeye ait.
$(d_1 < P) \text{ ve } (d_2 > P)$	O anki piksel üsttekinden farklı, fakat soldaki piksel ile benzer, demek ki soldaki piksel ile aynı bölgeye ait.
$(d_1 > P) \text{ ve } (d_2 < P)$	O anki piksel soldakinden farklı, fakat üstteki piksel ile benzer, demek ki üstteki piksel ile aynı bölgeye ait.
$(d_1 < P) \text{ ve } (d_2 < P)$	Soldaki ve üstteki piksellerin ikisiyle de o anki piksel benzer, bu yüzden hepsi aynı bölgeye ait.

Resimlerdeki bölgeleri tutmak için iki boyutlu bir dizi kullanılmıştır. Dizinin boyutu, resmin boyutuna eşittir. Her pikselin ait olduğu bölgenin numarası bu dizide tutulmaktadır. Başka bir deyişle, pikselin x,y olarak koordinatları kullanılarak ulaşılan dizideki “d[x][y]” sayısı, o pikselin ait olduğu bölge numarasını vermektedir.

Daha sonra resim soldan sağa ve aşağıdan yukarı olmak üzere iki “for” döngüsüyle taranır. Tarama sırasında yukarıda verilmiş olan dört durum sırasıyla kontrol edilir. Dördüncü adım için özel bir durum vardır. Bu adımda o anki piksel, solundaki ve yukarisındaki piksellerin her ikisiyle de aynı bölgede yer aldığı tespit edilmiş olur. Fakat, soldaki ve üstteki pikseller daha önceden farklı bölgeler olarak kaydedilmiş olabileceğinden bu durumun düzeltilmesi gerekir. Bu nedenle eş bölgelerde olan pikseller için ayrı bir dizi tutulmaktadır.

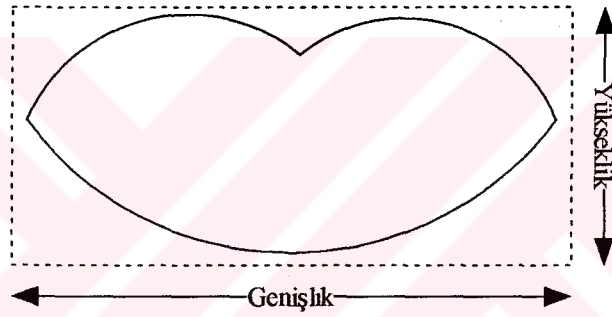
Sonuç olarak, dudak şekilde görüldüğü gibi bir dikdörtgen ile işaretlenmiş olur. Dudak biçimi bir dikdörtgen içine alınmış olduğundan, dudak biçiminin dört sayısal değeri elde edilebilir [Petajan, 1984]:

- Yükseklik
- Genişlik

- Çevre
- Alan

Bu değerler piksel sayısı olarak tutulmaktadır ve belli aralıklarla hesaplanarak, değerlendirilmeye alınır. Ayrıca bu değerler dışında; dudağın açık olup olmadığı, diş veya dilin görünüp görünmediği hakkında bir değer saptanmaktadır.

- Dudaklar açık
- Dudaklar kapalı
- Dişler görünüyor
- Dil görünüyor



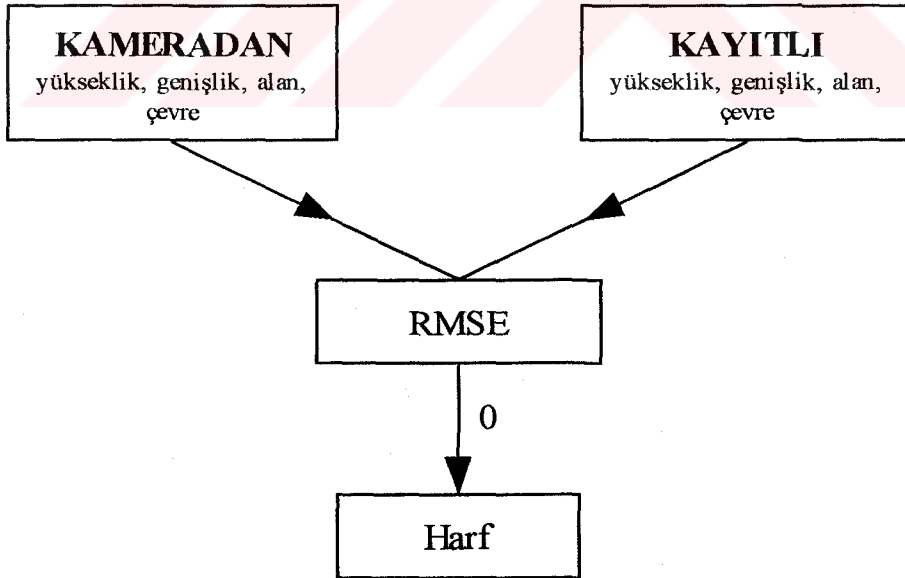
Şekil 5.2. Dudak Biçiminin Sayısal İfadesi

6. DEĞERLENDİRME

Dudak biçimini ifade eden dört değer, öncelikle önceden kaydedilmiş olan fonem-vaysem eşleşim tablosunda taranır. Kameradan alınacak olan görüntüdeki ufak tefek değişimlerden mümkün olduğunca az etkilenmek için, gelen değerleri ile önceden belirlenmiş değerler arasında “Root Mean Squared Error” işlemi yapılır.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}$$

RMSE değeri sıfıra ne kadar yakınsa, bu iki değer de birbirine o kadar yakındır. Kameradan gelen görüntülerden kaynaklanabilecek bir diğer problem ise ölçeklenme (scaling) problemidir. Kişinin kameraya uzaklığı veya yakınlığı, RMSE denkleminin saptayamayacağı kadar farklılaşabilir. Bu durumda elde edilen değerler kayıtlı görüntünün oranına göre tekrar ölçeklendirilir (rescaling).



Şekil 6.1. Verilerin Harflere Dönüşümü

Bu harfler biriktirildikten sonra, konuşmada olabilecek farklılıklardan kurtulmak için, dinamik zaman sarınımı (dynamic time warping) [Rabiner et. al.,

1993] tekniği kullanılır. Bir kelime “ALLLLTIII” veya “ALTIIIII” gibi farklı şekillerde söylenebilir. Görüldüğü gibi aynı kelimeler söylenmiş olmasına rağmen “T” harfinin söylendiği zaman farklıdır. “Dinamik zaman sarınımı” sayesinde gelen verilerin zamana bağlı olan farklılıklarından etkilenme miktarı azaltılmış olunur. Aşağıda dinamik zaman sarınımı yönteminin algoritması verilmiştir:

```
int DTWDistance(char s[1..n], char t[1..m], int d[1..n,1..m]) {
    declare int DTW[0..n,0..m]
    declare int i, j, cost

    for i := 1 to m
        DTW[0,i] := infinity
    for i := 1 to n
        DTW[i,0] := infinity
        DTW[0,0] := 0

    for i := 1 to n
        for j := 1 to m
            cost:= d[s[i],t[j]]
            DTW[i,j] := minimum(DTW[i-1,j] + cost,           // insertion
                                DTW[i,j-1] + cost)           // deletion
                                DTW[i-1,j-1] + cost)           // match

    return DTW[n,m]
}
```

Sistemin nasıl çalıştığını örneklemeye çalışırsak: “ALTI” kelimesinde harflerden oluşturduğumuz tabloda olanlar “A” ve “I” vardır. Bu yüzden önce dudakları kelimenin başında “A”, sonunda “I”nın tanımlandığı gibi biçimleniyorsa, büyük bir olasılıkla söylenen kelime “ALTI”dır.

Önceden kaydedilmiş 25 kelimeyle ile sistem test edilmiştir. Test iyi aydınlatılmış bir ortamda tek bir kişi (erkek) ile yapılmıştır. Konuşmacı, test sırasında mümkün olduğunca sabit bir konumda durmuş ve yavaş bir şekilde konuşmuştur.

Testte kullanılan kelimelerden başarılı olarak saptanabilenler şöyledir: Altı, Alper, Edirne, Kapalı, Foto, Uçak, Lale, Ama, Fatih, Fal, Petek. Başarılı olmayan kelimeler ise: Patates, Galiba, Çabuk, Soru, Neden, Ödemiş, Ankara, Haber, Salıncak, Balina, Katmer, Susam, Bardak, Deniz, Gazete’dir.

7. İMPLEMENTASYON DETAYLARI

320x240 çözünürlükte saniyede 25 kare yakalayabilen bir kamera ile çalışılmıştır. Programlama dili olarak kolaylığı ve işletim sisteminden bağımsızlığı nedeniyle Java seçilmiştir. Java programlama dilinin kameraya ulaşım ve kameradan kare kare resimleri alabilmek için hazırlanmış “Java Media Framework” eklentisinden yararlanılmıştır.

Yazılımdaki belli başlı kısımlar aşağıda listelenmiştir:

- LipTracker
- LipTrackerGUI
- JImagePanel
- BoundingBox
- CameraController
- ImageProcessor
- LipParams
- DynamicTimeWarping

Yazılım öncelikle, öncelikle sistemdeki birincil olarak kayıtlı olan kameraya ulaşmaya çalışır. Başarıyla ulaşabiliyorsa, o anki görüntüyü “CameraController” kare kare kameradan almaya başlar. Gelen her kare, “JImagePanel” yardımıyla ekranda gösterilir. Soldaki gösterim orijinal görüntüdür, soldaki ise işlenmiş görüntüdür. “LipTrackerGUI” kullanıcı arabiriminin tanımlanmış olduğu kısımdır. “ImageProcessor” kareyi işleyerek dudağın biçimini ortaya çıkarır ve ortaya çıkan kısım bir “BoundingBox” ile kırmızı çerçeve içinde işaretlenir. Son olarak bu çıkan biçim “LipParams”e yerleştirilerek değerlendirilmeye alınır. Harfler elde edildikten sonra bu harf dizilerinin önceden kayıtlı kelimelere uyup uymadığı “DynamicTimeWarping” ile yapılır.

8. SONUÇ

Bu çalışmada öncelikle dudak okuma ve faydalı olabileceği noktalar hakkında genel bilgi verilmiştir. Daha sonra dudak okuma sistemi oluşturabilmek için gerekli olan temel adımlar listelenmiştir. Dudağın biçiminin çıkarımı için kullanılan dört farklı yol anlatılmış ve bunlardan genel olarak en iyi sonucu veren, bu çalışmada da kullanılan yöntemler olan "RDY'ye Dayalı" ve "Kırmızıyı İhmal Etme" yöntemleri hakkında daha fazla detaya inilmiştir. Daha sonra bu biçimin sayısal ifadesini elde edebilmek için kullanılan yöntem olan "imge bölgesi" yöntemi anlatılmıştır.

Türkçe için iki dudaktan okuyabilen işitme engelli tarafından hazırlanmış bir fonem-vaysem eşleşim çalışmasına yer verilmiş ve örnek bir kelime grubunun bu eşleşim çalışmasına göre nasıl ifade edilebileceğinden bahsedilmiştir. Son olarak, bu tablo ve dudak çıkarımdan elde edilen verilerin nasıl değerlendirilip, dudak hareketi-kelime eşleştirmesi yapılabileceğinin detaylarına yer verilmiştir.

Sonuç olarak; bu çalışmada Türkçe bir dudak okuma sisteminin basit yoldan nasıl oluşturulabileceği gösterilmiş, ve oluşturulan eşleşim çalışması temel alınarak tek bir konuşmacı ile yapılan testlere %40'lık bir başarımla elde edilmiştir. Seçilmiş olan 25 kelimeden, 15 tanesi başarısız, 10 tanesi ise başarılı sonuç vermiştir. Sonuçlar, bu çalışmanın imge bölgesi ve dinamik zaman sarınımları gibi temel yöntemleri kullanan, sadece görsel verileri işleyen, ve herhangi bir kısıtlama getirmeyen bir yöntem izlediğini göz önüne alınarak değerlendirilmelidir. "Snakes", "Hidden Markov Model" gibi daha yeni ve etkin yöntemlerin kullanımı ile daha başarılı sonuçlar alınması mümkündür.

Karşılaştırma yapmak gerekirse: Bu başarımla düşük seviyeli dudak çıkarımı yöntemini kullanan Wark ve ark. [Wark et. al., 1998] aldığı %58'den kötü, yüksek seviyeli dudak çıkarımı yöntemi kullanan Sanchez ve ark. [Sanchez et. al., 1997] aldığı %48'lik orana yakındır. Bu çalışma ile özellikle tek bir konuşmacıyı temel alması ve dinamik zaman sarınımları kullanmasıyla benzer Petajan'ın [Petajan, 1984] çalışmasında ise %50'lik bir başarı sağlanmıştır.

Çizelge 8.1 Karşılaştırma

Çalışma	Başarı Yüzdesi
Coşkun	% 40
Wark	% 58
Sanchez	% 48
Petajan	% 50

KAYNAKLAR

- [Bregler et. al., 1993] Bregler, C., Manke, S., Hild, H., and Waibel, A. (1993). Bimodal Sensor Integration on the Example of "Speech-Reading". *Proceedings of the IEEE International Conference on Neural Networks*, sayfa 667-671.
- [Bull, 2005] Bull, D. (2005). Colour Tracker. <http://www.uk-dave.com/projects/java/colourtracker.shtml>
- [Chiou et. al., 1997] Chiou, G., Hwang, J.-N. (1997). Lipreading from color video. *IEEE Transactions on Image Processing*, sayfa 1192-1195.
- [Coianiz et. al., 1996] Coianiz, T., Torresani, L., and Caprile, B. (1996). 2D Deformable Models for Visual Speech Analysis.
- [Dodd and Campbell, 1987] Dodd, B. and Campbell R., Editors (1987). *Hearing by Eye: The Psychology of Lip-Reading*. Lawrence Erlbaum Associates, Hillsdale NJ.
- [Gonzalez and Woods, 2001] Gonzalez, R. C., Woods, R. E. (2001). *Digital Image Processing*. Prentice Hall.
- [Kleeman, 1994] Kleeman, M. (1994). *I See What You Say*. Hearing Visions.
- [Lewis and Powers, 2001] Lewis, W. T., Powers, D. M. W. (2001). Lip Feature Extraction Using Red Exclusion. *Visulation 2000, Pan-Sydney Workshop on Visual Information Processing*.
- [Microsoft, 2005] Microsoft Corp. (2005). <http://www.microsoft.com/speech/default.msp>
- [Perez, 2003] Perez, J. F. G., Lukas K., Frangi, A. F. (2003). Low Resource Lip Finding and Tracking Algorithm for Embedded Devices. *Eurospeech 2003, Vol. 3.*, sayfa 2253-6.
- [Petajan, 1984] Petajan, E. D. (1984). Automatic Lipreading to Enhance Speech Recognition. *Proc. Global Telecommunications Conference*, Atlanta, GA, sayfa 265-272.
- [Rabiner et. al., 1993] Rabiner, L., Juang B.-H. (1993). *Fundamentals of Speech Recognition*. Prentice Hall.
- [SAMPA, 2002] <http://www.phon.ucl.ac.uk/home/sampa/home.htm>.
- [Sanchez et. al., 1997] Ramos Sanchez, M. U., Matas J., and Kittler J. (1997). Statistical chromaticity models for lip tracking B-splines. *In Int. Conf. Audio and Video based Biometric Person Authentication, Crans Montana, Switzerland, 1997*.

[Steifelhagen et. al., 1997] Steifelhagen, R., Yang, J., and Meier, U. (1997). Realtime Lip Tracking for Lipreading. *Proceedings of Eurospeech 1997*.

[Türk ve Arslan, 2004] Türk, O., ve Arslan, L. (2004). Konuşma Terapisine Yönelik Otomatik Konuşma Tanıma Sistemleri. *II. Ulusal Dil ve Konuşma Bozuklukları Kongresi*.

[UCSC, 1996] University of California at Santa Claus (1996). Speech SciNotes. <http://scicom.ucsc.edu/SciNotes/9601/Speech/Lips.html>

[Visser, 1999] Visser, M., Poel, M., Nijholt, A. (1999). Classifying Visemes for Automatic Lipreading. *TSD*, sayfa 349-352.

[Wark et. al., 1998] Wark, T., Sridharan, S., and Chandran, V. (1998). An Approach to Statistical Lip Modelling for Speaker Identification via Chromatic Feature Extraction. *Proceedings of the IEEE International Conference on Pattern Recognition*, sayfa 123-125.

[Wikipedia, 2005] Wikipedia (2005). Lip reading. http://en.wikipedia.org/wiki/Lip_reading



ÖZGEÇMİŞ

Alper Coşkun, 1979 yılında İzmir’de doğmuş, aynı ilde ilk ve ortaokulu bitirdikten sonra; Mimar Sinan Teknik Lisesi Bilgisayar bölümünü tamamlamıştır. 2001 yılında İzmir’de yer alan Dokuz Eylül Üniversitesi’nin Bilgisayar Mühendisliği bölümünden mezun olmuştur.

