

# Bir Türkçe konuşma tanıma sisteminin anatomisi

## The anatomy of a Turkish speech recognition system

Cem Aksoylar, Süha Orhun Mutluergil, Hakan Erdoğan

Mühendislik ve Doğa Bilimleri Fakültesi

Sabancı Üniversitesi

{caksoylar,suhaorhun}@su.sabanciuniv.edu, haerdogan@sabanciuniv.edu

### Özetçe

*Bu bildiri geliştirdiğimiz Türkçe konuşma tanıma sisteminin yapısından bahsedeceğiz. Sistemin inşaa edilişii ve daha sonra sistem üzerinde yaptığımız deneyler hakkında bilgi vereceğiz. Sistem eğitimi için SUVoice veritabanı ile METU 1.0 veritabanını birlikte kullandık. Sınırlı dağarcıklı ve geniş dağarcıklı tanıma deneyleri yaptık. Modern saklı Markov modeli tabanlı konuşma tanıma sistemlerinin Türkçe için değişik sınaama koşullarındaki başarımını gösterdik. Basit sınaamalarda kelime hata oranı %1 civarında olurken geniş dağarcıklı sınaamalarda daha yüksek hata oranları elde ettik. Bu çalışma Türkçe konuşma tanıma konusunda daha ileri düzeyde çalışmalara bir temel teşkil edecek ve bu konudaki bilgi birikimine katkıda bulunacaktır.*

### Abstract

*In this paper, we present a Turkish speech recognition system we have developed. We will be giving information about the building of our system and tests we conducted on it. SUVoice voice database, along with METU 1.0, was used in the training of our acoustic models. We performed limited vocabulary and large vocabulary recognition tests. We have shown the performance of modern hidden Markov model based systems for Turkish speech recognition. For simple tasks, we can obtain 1% word error rate, whereas for large vocabulary tests the error rate is higher. This work will constitute a basis for more advanced future works in Turkish speech recognition and contribute to the accumulated knowledge in the field.*

### 1. Giriş

Ses tanıma sistemi genel olarak konuşma verisini yazıya çeviren bir yazılım olarak görülebilir. Bu tür bir sistemin gelişmesi insan makine etkileşimini daha kolay ve verimli kılacaktır. Yalnız, bir ses tanıma sisteminin geliştirilmesi zorlu bir süreç içerir. Her dil için farklı ses tanıma yazılımları oluşturulmak zorundadır, dolayısıyla her farklı dilde ses tanıma yazılımı o dile has özelliklerden kaynaklanan sorunları çözmek zorundadır.

Bütün bu zorluklara rağmen, etkili bir ses tanıma sistemi birçok alanda büyük faydalar sağlayabilir. Örneğin, birçok şirket çağrı merkezlerinde görevliler bulundurmaktadırlar ve bu da para ve insan gücü kaybına sebep olmaktadır. Bu kaybın önüne çağrı merkezindeki görevlilerin işini yapacak bir ses tanıma sistemi ile geçilebilir.

Ayrıca ses tanıma sistemleri metin dikte ettirmek için de kullanılabilir. Herhangi bir editör programına sesle yazı dikte

ettirmek, bu iş için harcanan süre ve emeği azaltacaktır. Ek olarak, doğrudan ses dosyalarından yazıya geçiş yapılabılır ve duyma güçlüğü çeken kişilerin yayımları takip etmesine olanak sağlar.

Gelişmiş bir ses tanıma sistemi, makinelere ses ile komut vermeye olanak sağlar. Teknolojik gelişmelerle televizyonlar, arabalar ve diğer elektronik aletler hiçbir aracı alet olmadan sesle yönetilebilir.

Ses tanıma sistemleri, gelişmiş çeviri yazılımlarıyla birlikte kullanılarak da faydalı olabilir. Birleştirilmiş bir sistemle farklı ana dile sahip iki kişi ortak bir dil bilmek zorunluluğu olmadan iletişim kurabilirler ve bu türden bir aygıt uluslararası konferanslarda simultane çeviri yaparak faydalı olabilir.

Bu bildiri geliştirdiğimiz Türkçe konuşma sisteminin bina edilmesini ve sınaanmasını anlatacağız ve bu sistemden elde ettiğimiz sınaama sonuçlarını tartışacağız. Umuyoruz ki, bu çalışmadaki bilgiler diğer Türkçe konuşma tanıma araştırmalarında yol gösterici olabilir.

Bildirinin ikinci bölümünde genel olarak konuşma tanıma sistemlerinin sahip olduğu yapıyı, özelliklerini ve kullandıkları yöntemleri özetleyeceğiz. Üçüncü bölümde, sistemin eğitilmesi için kullandığımız verinin yapısını anlatacağız. Dördüncü bölüme bizim deneylerimizi yaparken kullandığımız sistemi açıklayarak başlayacağız. Bunun için ikinci bölümde anlattığımız genel konuşma tanıma sistemi özelliklerinin bizim sistemimizde nasıl uygulandığını açıklayacağız. Daha sonra deneylerden elde ettiğimiz sonuçları göstereceğiz ve bu bölümün sonunda da sonuçlarımızdan elde ettiğimiz çıkarımları anlatacağız.

### 2. Konuşma Tanıma Problemi

Temel olarak konuşma tanıma sisteminin yaptığı görev bir konuşma verisini almak ve ne söylenildiğini tahmin etmektir. Yani sistemin girdisi konuşma verisidir ve çıktısı da tahmin edilen cümledir (hipotez cümlesi). Sistemin tahmin yapabilmesi için iki ana bölüme ihtiyacı vardır: öznitelik çıkartıcı ve dil çözümleyici. Öznitelik çıkartıcı, gelen ses dalgayapısındaki veriyi akustik öznitelik dizilerine dönüştürür ve dil çözümleyicinin kullanabileceği bir yapı oluşturur (bu araç hakkında daha detaylı bilgi bir sonraki bölümde mevcuttur). Daha sonra, dil çözümleyici bu öznitelikleri çıkartılmış diziyi alır ve cümleyi tahmin etmeye çalışır. Tahmin ederken iki temel model kullanılır: Dil modeli ve akustik model. Bu modeller hakkında da detaylı bilgi bu bölüm içinde verilecektir. Temel olarak tahmin işlemi aşağıdaki Bayes denklemi sayesinde gerçekleştirilir:

$$\hat{W} = \arg \max_W P(W | A) = \arg \max_W P(A | W)P(W) \quad (1)$$

Burada  $\hat{W}$  hipotez cümlesidir. Bu cümle eşitliğin ikinci tarafında da görülebileceği gibi olası cümleler içinden akustik diziyile uyuma ihtimali en yüksek olan cümledir.  $P(W|A)$  verilen akustik veriye göre ( $A$ ) mevcut cümlelerin ( $W$ ) akustik diziyile eşleşme olasılığını simgeler. Eşitliğin ikinci kısmı doğrudan Bayes kuralının uygulanması ile elde edilir. Burada  $P(W)$ ,  $W$  cümlesinin oluşma olasılığıdır ve dil modeline göre hesaplanır.  $P(A|W)$  verilen cümleye göre bizdeki akustik dizinin oluşması olasılığıdır ki bu da akustik modelle hesaplanır. Bayes kuralından gelen  $P(A)$   $W$ 'dan bağımsız olduğu için sadeleştirilmiştir.

### 2.1. Öznitelik Çıkarıcı

Daha önce de bahsedildiği gibi ses tanıma işleminden önce ses verisi bir dizi parametre vektörüne dönüştürülmelidir. Bu sebeple öznitelik çıkarıcı ses verisini zaman açısından birbirleriyle keşimeleri boş olmayan kesitlere ayırır ve her bir kesitin özniteliklerini çıkarır.

Öznitelik çıkarımı için birçok farklı yöntem bulunsa da insan kulağının algılamasına benzer frekans çözümlemesi sunduğu için filter-bank analizi daha yaygındır. Bu avantajından dolayı bizim de ses tanıma için kullandığımız HTK programı filter-bank yaklaşımını adapte eder [1]. Eşit frekans aralıklı üçgen filtreler oluşturulur ve ses dalgasının Fourier dönüşümü bu filtrelerle çarpılarak toplanır. Böylelikle filtrelerdeki değerlere göre ağırlıklı bir toplam elde edilir. Yeni oluşturulan bu veriler ve MFCC (Mel-frequency cepstral coefficients) kullanılarak vektör öznitelikleri çıkarılır.

### 2.2. Akustik Model

Ses tanıma sistemlerinin akustik model temelini Saklı Markov Modelleri (Hidden Markov Models- HMM) oluşturur. En küçük akustik birim fonemdir ve fonem bazlı modellerde genellikle 3 adet HMM durumundan oluşur.

HMM'in model parametreleri durum değiştirme olasılıkları (state transition probabilities -  $a_{ij}$ ) durum gözlenme yoğunlukları (state observation density -  $b_j(o)$ ) ve başlangıçtaki durum dağılımlarıdır (initial state distributions -  $\Pi_j$ ).  $b_j(o)$  'nin hesaplanması Gauss karışım dağılımı olasılık fonksiyonlarının kullanılmasıyla hesaplanır. Kullanılabilecek Gauss karışımı sayısı artırılabilir.

Monofon kullanımı etrafındaki diğer seslerden bağımsız bir ses tanıma sistemi sağladığı için ve bir fonem birden çok sese karşılık gelebileceği için monofon tanıma sistemi etkili bir çözüm sunmaz. Trifonlar ise her fonemin sağında ve solunda bulunan fonemler ile birlikte modellenmesidir, bu sebeple trifonlar kullanmak akustik değişkenliği düşürür ve daha etkili bir tanıma sağlar.

### 2.3. Dil Modeli ve Gramer

Dil modelleri veya gramerler, konuşma tanıma yapılırken, söylenen kelimelerin akustik tanımının yanında, içinde buldukları bağlamlara göre de tanınmalarına olanak verirler.

Gramerler elle belirlenmiş cümle kuralları tanımlarlar, akustik model bir kelimenin olasılığını en yüksek olarak belirlese de, bu kelime gramer kurallarının dışında ise tanıma olasılığı

olmaz, sadece bu kurallarla belirlenmiş kelimeler tanınabilir. Bu nedenle gramerler sadece sınırlı dağarcıklı konuşma tanıma sistemlerinde kullanılırdılar. Bu tür elle belirlenen gramerlerin yazımında da BNF gösteriminden yararlanılır [2].

Öte yanda dil modelleri, eğitim metinlerinden eğitilir ve cümlelerin o dildeki olasılıklarını oluşturur. Eğer cümlelerin kelimelerden oluştuğunu varsayarsak,  $w_i$  kelimelerinden oluşan  $W$  cümlesinin tanıma olasılığı aşağıdaki gibidir:

$$P(W) = P(w_1, w_2, \dots, w_n) = \prod_{i=1}^n P(w_i | w_1, \dots, w_{i-1}) \quad (2)$$

Dil modelleri uygulamalarında genellikle N-gram modeller kullanılır, bu modeller de denklemin sağ tarafında sadece  $w_i$ 'den önceki N-1 kelimeyi kullanarak denkleme yaklaşır. Böylece yukarıdaki denklem aşağıdaki denkleme dönüşmüş olur:

$$P(W) \approx \prod_{i=1}^n P(w_i | w_{i-N+1}, \dots, w_{i-1}) \quad (3)$$

Bu N-gram olasılıkları ise aşağıdaki formülle hesaplanır:

$$P(w_i | w_{i-N+1}, \dots, w_{i-1}) = \frac{C(w_{i-N+1}, \dots, w_{i-1}, w_i)}{C(w_{i-N+1}, \dots, w_{i-1})} \quad (4)$$

Buradaki  $C(w_{i-N+1}, \dots, w_{i-1}, w_i)$  belirtilen N kelimenin eğitim metinlerindeki geçme sayısını gösterirken,  $C(w_{i-N+1}, \dots, w_{i-1})$  ise belirtilen N-1 kelimenin metinlerde geçme sayısını gösterir. N=2 durumunda ise bigram dil modeli elde edilir.

Eğitim verisinde görülmeyen kelimelerin tanınmaları için de dil modeli içerisinde düzleme (smoothing) yöntemi uygulanarak onlara da birer olasılık değeri atanır.

Bu dil modellerinde oluşturulan istatistiksel dağılımlara göre kelimelerin tanıma olasılıkları, aynı akustik modelde olduğu gibi, değişim gösterir.

### 2.4. Değerlendirme Ölçütleri

Konuşma tanıma sisteminin performansını ölçmek için birkaç bilinen ölçüm yöntemi mevcuttur. En doğru kıstas hipotez cümlesi ile referans cümle arasındaki farkların değerlendirilmesidir. Bu sebeple kelime hata oranı (Word Error Rate –WER) aşağıdaki gibi hesaplanabilir:

$$WER = \frac{D + S + I}{N} \times 100 \quad (5)$$

Burada N referans cümledeki toplam sembol (label) sayısını, D referans cümleye göre hipotezdeki silinen sembol sayısını, S değişen sembol sayısını, I da fazladan eklenmiş sembol sayısını temsil etmektedir. WER'yi bilmek bize doğru bilinen kelime oranını (percent accuracy) da sağlar.

Bunun dışında Kök hata oranı (Stem error rate – SER) ve Harf hata oranı (Letter error rate – LER) da değerlendirme ölçütü olarak tanımlanabilir. SER, referans ve hipotez cümlelerin kelimelerinin eklerinin atıldıktan sonra, köklerde WER'nin hesaplanmasıyla bulunabilir. LER ise harfleri tek tek

ayırdıktan sonra harfler arasında WER'nin bulunmasıyla hesaplanır.

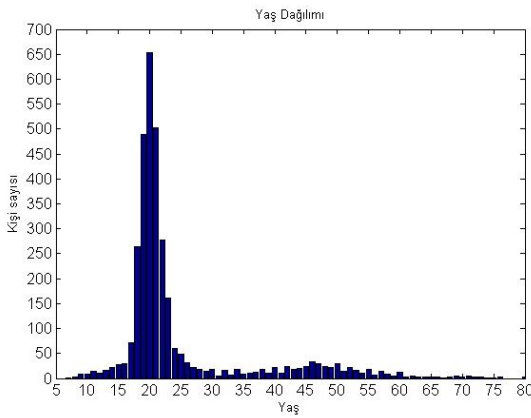
### 3. Veritabanı

Akustik modellerin eğitiminde temel olarak SUVoice ve Metu 1.0 veritabanları kullanıldı.

#### 3.1. SUVoice Veritabanı

Sabancı Üniversitesi öğrencileri tarafından bir proje dersi kapsamında her dönem düzenli olarak toplanan, 6 yıllık bir konuşma veritabanıdır. Toplam 3444 okuyucunun ses kayıtlarından oluşturulmuştur. Yaklaşık 70 saati sessizlik olan 185 saatlik konuşma verisi içerir. 46 farklı metin dosyasından 4087 özgün cümlelerin okunmasıyla oluşturulmuştur.

Veritabanı genelde üniversite içinden toplandığı için çoğunlukla 18-25 yaş grubundaki konuşmacıların verisini içerir. Konuşmacıların yaş dağılımı Şekil 1'de gösterilmiştir.



Şekil 1: Yaş dağılımı histogramı.

Konuşmacıların 2055'i erkek, 1389'u ise bayandır. Ayrıca 2447 adet sigara içmeyen, 997 adet aktif olarak sigara içen denek kullanılmıştır.

#### 3.2. METU 1.0 Veritabanı

Orta Doğu Teknik Üniversitesi ve University of Colorado at Boulder'ın ortak çalışması olan SONIC konuşma tanıma sisteminin Türkçe'ye uyarlanması projesi için ODTÜ'de toplanmış verilerden oluşmaktadır. Yaklaşık 500 dakikalık ses verisi mevcuttur. Her konuşmacı yaklaşık 40 cümle okumuştur ve 2462 özgün cümleden oluşmaktadır. Veriler 68'i erkek 52'si kadın 120 kişinin ses kayıtlarını içerir [3].

### 4. Deneyler ve Tartışma

Konuşma tanıma sisteminin bina edilmesi ve sınavında HTK programı kullanıldı. Programın eğitimi için elimizdeki veritabanı eğitim verisi ve sınav verisi olarak ikiye ayrıldı. Sınav verisi eğitim verisi olarak kullanılmadı.

Deney sonuçları ise üç farklı şekilde 2.4 bölümünde anlatıldığı gibi WER, SER ve LER olarak ölçüldü. Bu sonuçlarla ilgili daha detaylı karşılaştırma ise 4.4'de açıklanacak.

#### 4.1. Konuşma Tanıyıcı Yapısı

Konuşma tanıyıcı olarak kullandığımız HTK eğitimi için temel olarak HMM kullanılmaktadır. Biz de eğitim için 3 durumlu

HMM'ler kullandık. Öznitelik çıkarımları da MFCC yöntemine göre yapıldı. Bu yöntemle çıkarılan vektörlerin sonuna Enerji seviyesi (energy), fark (delta) ve ivmelenme (acceleration) vektörleri eklendi ve öznitelik ortalama çıkartımı (cepstral mean normalization) yapıldı.

Akustik modelin eğitimi için Türkçe'deki 29 harfe karşılık gelen 29 fonem kullanıldı. Bunlara ilaveten kelimeler arasındaki kısa boşlukları yakalamak için kısa durak ve cümle aralarındaki boşlukları yakalamak için sessizlik modelleri eğitildi. Bunların yanında yabancı bazı harflerin fonemleri de önlem olarak oluşturuldu. Toplamda 34 fonem elde edildi. Elde edilen fonem seti MetuBET alfabesinin bir alt kümesini teşkil etmektedir [3]. Her bir model için 3 durum (state) oluşturuldu ve her bir durum için 12 Gauss karışımı kullanıldı.

Öncelikle bu 34 fonem üzerinden monofon bir eğitim gerçekleştirildi. Daha sonra mevcut fonemler kullanılarak trifonlar oluşturuldu. Bütün 3'lü fonem kombinasyonlarını eğitmeye yetecek kadar veri olmadığı düşünüldüğünden bağlı durum (tied-state) trifonlar kullanıldı. Bağlı durumların oluşturulması için Türkçe için uygun bir karar ağacı (decision-tree) kullanıldı. Toplamda 4167 trifon modeli ve 50004 Gaussian mixture elde edildi.

Dil modeli olarak ise Osman Büyük'ün "*Sub-Word Language Modeling for Turkish Speech Recognition*" isimli yüksek lisans tezinde kullandığı kelime tabanlı bi-gram dil modelini kullandık. [4][5].

Sınama deneyleri Viterbi algoritması kullanılarak yapıldı, bu konuda değinilmesi gereken bir nokta da, kullanılan sınav parametreleridir. Özellikle sınav sonuçlarını etkileyen parametrelerden biri dil modeli logaritmik ağırlığıdır. Bu dil modeli ağırlığı değerleri (-s) için birden fazla deney yapılmış, 5, 7, 10 ve 15 ile elde edilen değerler 4.4'te karşılaştırılmıştır.

Sınavı etkileyen diğer önemli parametrelerden kelime ekleme logaritmik olasılığı olarak 0, budama (pruning) parametresi olarak da 120 kullanıldı.

#### 4.2. Sınama Verisi

Sınama deneyleri, sınırlı dağarcıklı ve geniş dağarcıklı olmak üzere iki ayrı kategoride yapıldı. İki kategorinin deneylerinde de SUVoice veritabanının eğitime dahil edilmeyen alt kümeleri kullanıldı.

Sınırlı dağarcıklı tanıma deneylerinde de iki ayrı deney metni türü kullanıldı; bunlardan ilki, ard arda rastgele dört rakam okunan deneylerdi, diğeri ise toplam 160 yer ismi (Türkiye illeri, ilçeleri vb.) arasından birinin okunduğu deneylerdi. Rakam sınav deneyleri için 54 kişiden toplam 5397 ses dosyası kullanılırken, yer ismi deneyleri için ise 14 kişiden 2238 ses dosyası kullanıldı.

Geniş dağarcıklı tanıma deneylerinde de iki farklı metin türü kullanıldı; bunların da ilki Yaşar Kemal'in romanı İnce Memed'den alınan cümleleri içerir, ikincisi ise spor haberleri arasından toplanmış cümlelerden oluşur. İnce Memed verisi 88 özgün cümle içermekte olup, sınav için 18 kişiden toplanan 1905 ses dosyası kullanılmıştır. Spor haberleri verisi de aynı şekilde 88 özgün cümle içerir, ve 16 kişiden toplanan 1408 ses dosyası ile sınav yapılmıştır. Bu sınav verileri [4]'te kullanılan sınav verilerinin aynısıdır.

Geniş dağarcıklı tanıma deneyleri için 30132 kelimededen oluşan bir sözlük kullanılmıştır.

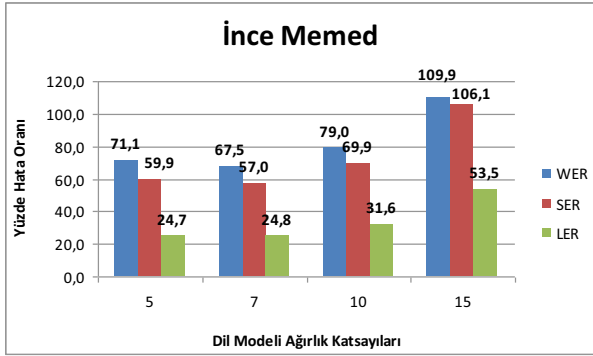
### 4.3. Deney sonuçları

Sınırlı dağarcıklı deneylerde yaptığımız yer tanıma ve rakam tanıma deney sonuçları Tablo 3'te gösterilmiştir. Rakamlarda cümle tanıma oranını farklı olmasının sebebi rakam sınamasında bir cümlede 4 rakam okunmasıdır. Yer isimlerinde iki hata oranının da aynı olmasının sebebi ise her cümlenin sadece bir yer ismini içermesidir.

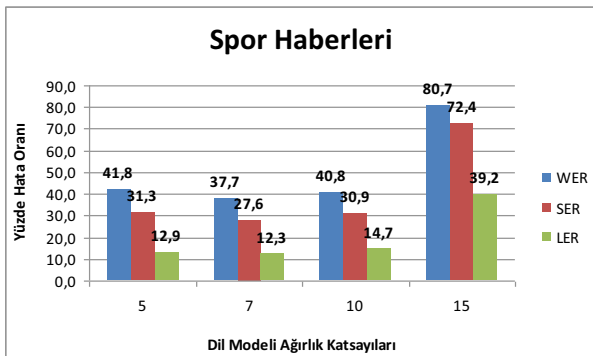
Tablo 3: Sınırlı dağarcıklı deney sonuçları

	Kelime hata oranı	Cümle hata oranı
<b>Rakamlar</b>	%1.0	%3.5
<b>Yer isimleri</b>	%1.2	%1.2

Geniş dağarcıklı deneyde ise İnce Memed ve spor haberleri metinleri sılandı. Deney sonuçları Şekil 2 ve Şekil 3'te görülebilir. Şekillerde yeşil sütunlar harf hata oranını (LER), kırmızı sütunlar kök hata oranını (SER) ve mavi sütunlar ise kelime hata oranını (WER) göstermek-tedirler.



Şekil 2: İnce Memed roman metni sınama sonuçları



Şekil 3: Spor haberleri metni sınama sonuçları

### 4.4. Tartışma

Deney sonuçlarına bakıldığında, beklenildiği gibi sınırlı dağarcıklı deneydeki hata yüzdesi, geniş dağarcıklı deneylere göre çok daha düşüktür. Çünkü sınırlı dağarcıklı deneydeki olası sözcük sayısı ve sözcüklerin birbirlerine benzerlikleri daha azdır. Böylece tanıma sistemi daha az seçenek arasından daha isabetli seçimler yapabilmektedir. Geniş dağarcıklı deneyde, dil modelinin tanımayla yaptığı katkılara rağmen olası sözcük

sayısının fazlalığı nedeniyle, deney sonuçları sınırlı dağarcıklı deneydeki isabet oranına yaklaşmamaktadır.

Bir diğer gözlem de iki geniş dağarcıklı sınama sonuçları arasındaki hata oranları farkıdır. Bütün ölçütlerde (SER, WER, LER) Spor haberlerinin hata oranı İnce Memed'e göre daha düşüktür. Bunun nedenlerinden biri İnce Memed'te geçen kelimelerin spor haberlerindekiyle göre, akustik modelin ve dil modelinin eğitim verisinde daha az bulunmasıdır. Bu nedenlerin en önemlisi de, romandaki - sistem tarafından tanınması imkansız olan - sözlük dışı kelimelerin sayısının (out of vocabulary) fazla olması ve dolayısıyla doğru tanıma yüzdesinin düşmesidir. Bu sonuçlar [4]'teki sonuçlarla karşılaştırıldığında spor haberi sınama verisi için %46.4 WER değerinden %37'ye düştüğü görülmektedir. Bunun sebebi daha fazla veriden daha iyi bir akustik modelin eğitilmiş olmasıdır.

Geniş dağarcıklı deneylerdeki doğruluk oranını değiştiren bir diğer gözlemlenebilir ölçüt de, s (dil modeli ağırlığı) parametresidir. Şekil 2 ve Şekil 3'te s değerinin değişiminin etkileri görülebilir. Burada gözlemlenebilecek ilk şey s değerinin değişmesinin bütün ölçütlerde (SER, WER, LER) aynı yönde (artma veya azalma) etki yapmasıdır. Ayrıca bütün ölçütler için her iki deneyin de tanıma oranını en fazla yapan değer 7'dir. s değerinin 7'ye göre artması veya azalması doğruluk oranının düşmesine yol açmaktadır.

## 5. Sonuçlar

Bu bildiri geliştirdiğimiz Türkçe konuşma tanıma sisteminin yapısından ve işleyişinden bahsettik. Bu sistem üzerinde yaptığımız deneylerin sonuçlarını açıkladık ve çıkan sonuçları karşılaştırdık. Sistemimiz sınırlı dağarcıklı konuşma tanıma iyi sonuçlar vermesine rağmen geniş dağarcıklı tanıma düşük başarımla gösterdi. Bu sebeple, gelecek çalışmalarda geniş dağarcıklı tanıma oranının artması için dil modelinin geliştirilmesi düşünülebilir. Daha iyi bir dil modeli için istatistik olmayan bir gramer kullanılabilir. Ayrıca dil modelinin eğitimini kelimelerden yapmak yerine, kelime altı parçacıklardan yapılabilir.

Genel tanıma başarımının artması için de, tanıma programına vermeden önce, ses dosyası üzerinde etkili yazılımlar kullanılarak, gürültü azaltılması gibi işlemler yapılabilir.

## 6. Kaynakça

- [1] Young, S., et al., "The HTK Book (for HTK Version 3.4)," Cambridge University Engineering Department, 2006.
- [2] Linz, P., "An introduction to formal languages and automata," Jones and Bartlett Publishers, 1997.
- [3] Salor, Ö., Pellom, B., Çiloğlu, T., et al, "On developing new text and audio corpora and speech recognition tools for the Turkish language," presented at 7th International Conference on Spoken Language Processing, Denver, Colorado, USA, 2002.
- [4] Büyük, O., "Sub-word language modeling for Turkish speech recognition," M.S. thesis, Sabancı University, Istanbul, Turkey, 2005.
- [5] Erdogan, H., Buyuk, O., Oflazer, K., "Incorporating language constraints in sub-word based speech recognition," presented at IEEE Automatic Speech Recognition and Understanding Workshop, Cancun Mexico, 2005.