2007 Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi Sayı:21, s.79-97

HTK İLE KONUŞMACIDAN BAĞIMSIZ TÜRKÇE KONUŞMA TANIMA SİSTEMİ OLUŞTURMA

Nursel YALÇIN[1]
Ömer Faruk BAY[2]

ÖZET

Konuşma tanıma, bir mikrofon kullanılarak kullanıcının söylediklerini ekrana yazdıran bir teknolojidir. Bu makalede konuşmacıdan bağımsız Türkçe konuşma tanıma sistemi oluşturma aşamaları anlatılmış ve konuşma tanıma sisteminde kullanılan algoritma açıklanmıştır. Konuşma tanıma sistemi, hazırlanan metin editörünün içine entegre edilmiştir. KTM adındaki Konuşma Tanıyan Metin Editörü sesli komutlarla da kullanıcı tarafından kontrol edilebilmektedir. Konuşma tanıma için ses eğitimi HTK ile gerçekleştirilmiştir. Genel olarak geniş dağarcıklı bir sistem hedeflendiği için sözcük-altı akustik modeller kullanılmıştır.

Anahtar Kelimeler: Viterbi Algoritması, Geniş Dağarcıklı Konuşma Tanıma, Bütünleşik SMM

CREATION OF SPEAKER INDEPENDENT TURKISH SPEECH RECOGNITION SYTEMS WITH HTK

ABSTRACT

Speech recognition is a technology writes speech of user to screen that is used a microphone. In this paper is described constituted phases of speaker independent turkish speech recognition system and the algorithm, which is used in speech recognition system, is explained. This speech recognition system is integrated into the text editor. Speech Recognizer Text Editor is KTM, which can be controlled with voiced commands also by user. Acoustic sound training for speech recognition is realized with HTK(Hidden Markov Toolkit). In general this paper is used sub-word acoustic models because a large vocabulary speech recognition system is aimed.

Key Words: Viterbi Algorithm, Large Vocabulary Speech Recognition, Integrated HMM

1. GİRİŞ

Bazen rahatımızı düşünerek bilgisayarlardan zeki davranışlar göstermesini isteriz. İşte bu noktada bilgisayarların konuşma dilini anlamalarını bekleriz. Sesli bilgi girişi önemli ölçüde göz ve ellere düşen yükü azaltmakta ve uzaktan doğal iletişimi sağlamaktadır. Başlangıç aşamasında "anlam-metin" ve "metin-anlam" dönüşümleri göz ardı edilirse ilk iş olarak "metin-ses sinyali" ve "ses sinyali-metin" dönüşümlerinin yapılması gerekir. Bunlardan ilkine sesin analizi ikincisine de sesin sentezi denilmektedir.

Sesli veri girişi, klavyeden olana nazaran çok daha (yaklaşık 3-5 kez) hızlıdır. Ayrıca bu sistemler yabancı dil öğreniminde, uzaktan yönlendirme sistemlerinde ve tıpta büyük

[1] *Gazi Üniversitesi, Endüstriyel Sanatlar Eğitim Fakültesi, Bilgisayar Eğitimi Bölümü, Gölbaşı/Ankara,06830, nyalcin@gazi.edu.tr*
[2] *Gazi Üniversitesi, Teknik Eğitim Fakültesi, Elektronik ve Bilgisayar Eğitimi Bölümü, Teknikokullar/Ankara,06500, omerbay@gazi.edu.tr*

kolaylıklar sağlamaktadır. Körler için konuşan kitaplar, özürlüler için arabaların kullanımı ve sağırlar için konuşma görüntüsünün üretimi vb. bunlara örnek gösterilebilir (Nabiyev, 2005:691)

Konuşma tanıma, bir mikrofon kullanılarak kullanıcının söylediklerini ekrana yazdıran bir teknolojidir. Bundan dolayı konuşma tanıma özelliklerinin kullanılabilmesi için sistemde bir ses kartı ve buna bağlı bir mikrofon bulunması gerekmektedir. Konuşma tanıma sistemlerinin uygulama alanları oldukça fazladır. Ticari açıdan bakılacak olursa, konuşma tanıma potansiyel olarak çok büyük pazarı olan bir teknolojidir. Konuşma işleme teknolojisindeki gelişmeler sayesinde, ağır işitenler daha iyi işitebilmekte, sağır olanlar ise konuşmanın anında yazıya çevrilmesi ile canlı yayınlardaki konuşmaları anlayabilmektedirler (Öncül, 1993:1).

Konuşmacının sözcükler arasında 100 - 250 ms'lik duraksamalarla yaptığı konuşmayı tanımak için Ayrık Sözcük Tanıma sistemleri kullanılırken, hiçbir kısıtlama olmaksızın doğal ve 150 - 250 sözcük / dakika kadar hızlı konuşmak için Sürekli Konuşma Tanıma sistemleri kullanılır. Bir diğer şekli de sözcükler arasında durmaksızın her sözcüğün açık ve anlaşılır biçimde bastırılarak söylendiği konuşmayı tanıyan Bağlı Sözcük Tanıma sistemleridir (Gökhan, 1997:1). Konuşma Tanıma Sistemleri, Konuşmacıya Bağımlı ve Konuşmacıdan Bağımsız olarak da iki ayrı grupta değerlendirilir.

Konuşma tanıma da en büyük problemlerden biri zaman normalizasyonudur. Bir kelimenin sürekliliğinde ve zaman dağılımında farklılıklar bulunmaktadır. Bu farklılıklar, yalnızca değişik kişiler için değil, aynı kişiden farklı zamanlarda alınan örnekler için de geçerli olmaktadır (Nabiyev, 2005:710). Kütüphane bilgileri ile tanıma sistemine girdi olarak alınan kelime arasında zamansal sıraya koyma çok önemlidir. Zamansal sıraya koyma işleminde iki teknik sıkça kullanılmaktadır. Bunlardan ilki Dinamik Zaman Eşleştirme (Dynamic Time Warping - DTW) ses verisinin olasılıklara dayalı markov modelinin kullanımıdır (Nabiyev, 2005:710).

Dinamik zaman eşleştirme yönteminde zaman ekseninin bozulması söz konusudur. İki ayrı işlenecek örneğin (template) karşılık gelecek parçalarının zamansal dizilimi, birinden diğerine haritalamakla (mapping) yapılabilir (Nabiyev, 2005:710).

İkinci yaklaşım ise sese hem deterministik, stokastik olarak modellenebilen bir sinyal olarak bakmaktır. Ses üzerinde zaman bölgesinde yapılan çalışmalar genel olarak Saklı Markov modelini kullanırlar. Bu model markov zincirini esas alarak her kelime veya fonem için bir zincir üretir. Fizyolojik ses üretiminde ses organları sesi üretmek için çeşitli posizyonlara taşınır. Oluşan sesin değişikliğinin derecesi vardır. Bu yüzden mümkün çıkışların kümesini belirli bir durum ile ilişkilendirmek mümkündür. Kelime veya cümleyi söylerken ses organlarının bir durumdan başka duruma geçişi söz konusudur. Organın her posizyonuna bir durum atanırsa, durumlar arası geçişlerden bahsedilebilir. HMM de durumlar ve bu durumlar arası geçişlerin olasılık temsiline dayalıdır. Eğitim esnasında, belirtilen parametreler belirlenir (Nabiyev, 2005:711). Tanıma işleminde ise giriş ve taban bilgileri arasındaki benzerlik Bayes kuramı ile bulunur. Kısacası bu modelin temelinde istatistiksel değerlendirmeler sonucunda durum ve geçiş parametreleri bulunur. Bu parametreler tanıma sisteminin veri tabanında tutulur. Gözlemler durumun ihtimal fonksiyonu olduğu için Markov durumlar dolaylı olarak izlenmektedir. Bu sebeple yönteme Saklı (Hidden) nitelenmesi yapılmaktadır. Markov zincirlerinin geliştirilmiş şekli olan Saklı Markov Modeli, Baum ve meslektaşları tarafından 1970'li yılların başlarında

geliştirilmiştir. Yöntemin temelini (t-1) anındaki durum bilindiğinde t anındaki durumun değerlendirilmesi mantığı oluşturmaktadır. Konuşma tanımada istatistiksel işlem modeli tabanlı Markov zinciri günümüzde en başarılı ses modelidir. Bir durumdan bir sonraki duruma geçiş olasılığı yalnızca o andaki duruma bağlıdır. Önceki durumların ise geçiş olasılığına etkisi yoktur (Nabiyev, 2005:711).

Bu makalede konuşma tanıma için akustik ses eğitimi HTK (Hidden Markov Toolkit) denilen bir araç kiti tarafından yapılmıştır. HTK, ağırlıklı olarak konuşma tanıma sistemleri için gerekli olan Saklı Markov Model (SMM) lerin oluşturulması ve daha sonra oluşturulan SMM'ler ile tanıma ve değerlendirme için kullanılabilecek çeşitli araçların bir araya geldiği bir yazılım paketidir. Bunun yanında konuşma tanımada kullanılan farklı yöntemlerde vardır. Bunlardan ilki genellikle yapay zeka teknikleri içinde bulunan yapay sinir ağları, ikincisi ise dinamik zaman eşleştirme teknikleridir. KTM yazılımı ilköğretim birinci sınıflara ilkokuma yazma öğretimi için geliştirilmiş bir çalışmadır. Bu makalede KTM'nin ilköğretim alanında nasıl kullanıldığının açıklanması yapılmayacak ancak konuşma tanıma için oluşturulan algoritmalar verilecektir.

Makalenin ikinci bölümünde HTK araç kiti ile yapılan çalışmalara yer verilerek ses verilerinin eldesi, öznitelik verilerinin eldesi, teklises eğitimi ve bundan üçlüseslere geçişler, karar ağacı birleştirimi ve karışım sayılarının artırılmasıyla oluşturulan çalışma ortamı ve elde edilen dosyalarla ilgili çalışmalar anlatılmıştır. Üçüncü bölümde ise Konuşma Tanıma Çekirdeği ve Algoritmaların Tanıtımına yer verilerek tanıma işleminin gerçekleştirilmesi için gerekli diğer algoritmalar açıklanmıştır. Dördüncü bölümde ise oluşturulan yazılımın genel olarak kullanılmasına yönelik olarak konuşma tanıma özellikleri verilmiş beşinci bölümde ise yapılan çalışmaya yönelik sonuçlar belirtilmiştir.

2. HTK (HIDDEN MARKOV TOOLKIT) İLE YAPILAN ÇALIŞMALAR

Konuşma tanıma sistemimizde kullanılan akustik modellerin üretimi HTK adlı araç kiti sayesinde gerçekleştirilmiştir. Bu sistem konuşma tanıma işlemleri için gerekli olabilecek çeşitli işlevleri (özellik çıkarımı, eğitim, tanıma, değerlendirme) yerine getirebilecek birimlerden oluşmuştur. Bu çalışma ile ilgili ayrıntılar aşağıdaki gibidir:

2.1. Ses verileri

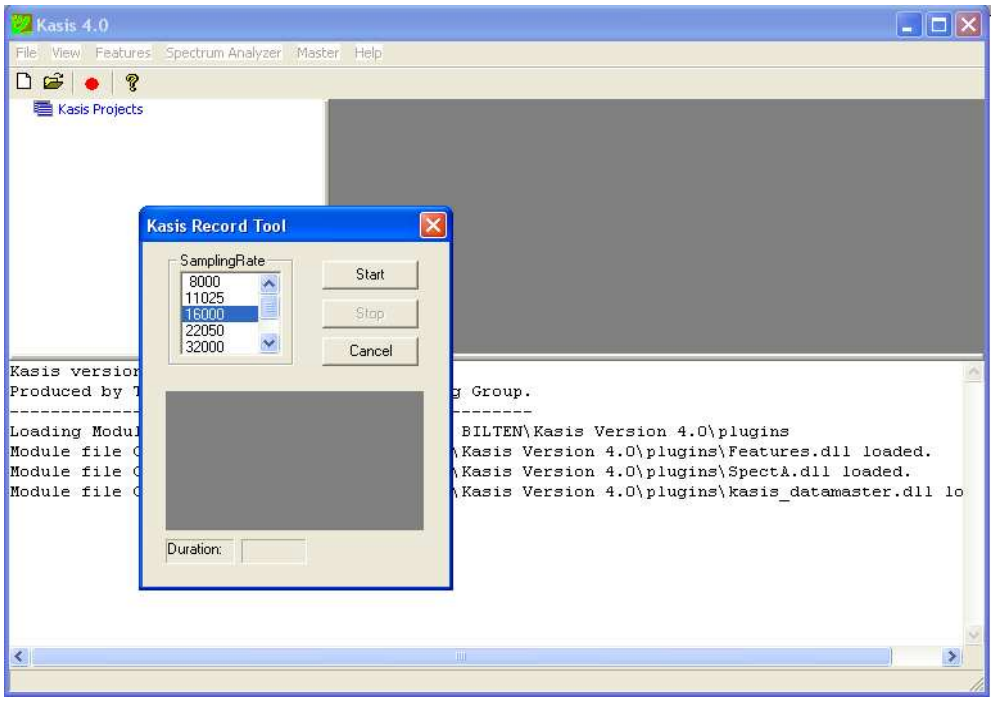
Model eğitiminde kullanılan ses verileri 16 kişiden alınan toplam 1600 adet cümleden oluşmaktadır. Her cümle ayrı bir ses dosyasına, içeriğine göre numaralandırılmış dosya isimleri ile okunmuştur. Sesler 16kHz örnekleme frekansında 16-bit mono olarak kaydedilmiştir. Cümleler sürekli konuşma olarak hazırlanmıştır. Sessizlik olan bölgeler virgül ile ayrılmıştır. Kullanılan cümlelerin seçimindeki kriter dengeli ses dağılımı ve bağlam dağılımının sağlanmasıdır. Ses verilerinin kaydedilmesinde kullanılan cümlelere örnekler;

"o, bütün yıl , senin elbiseni yağlı yıkama suyunda tuttu"
"bana yağlı bir paçavrayı bu şekilde tutmamı söyleme"
"bu bizim için kolaydı"
"jale çok sıkı çalışarak, daha fazla para kazanabilir"
"o benden daha zayıf"
"parlak güneş ışığı okyanus üzerinde parıldıyor"
"hiçbir şey suçsuzluk kadar hatır kırıcı değildir"
"aptal maddeler üzerinde üzülmek ya da avazı çıktığınca bağırmak niye"

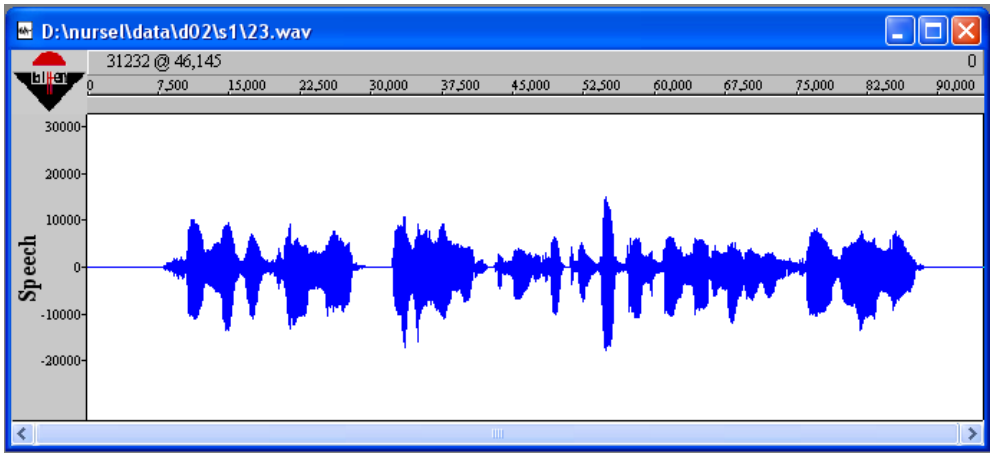
Tabiî ki dengeli ses dağılımına örnek olması açısından bu birkaç örnek yeterli değildir. Ses kaydında amaç dengeli ses dağılımını ve bağlam dağılımını oluşturmak olduğu için çok önemlidir. Ses veri tabanında kullanılan bu cümlelerdeki sesler gerekli olan bütün dengeli seslere karşılık gelebilecek şekilde olmalıdır. Yoksa sadece belli bir sözcük yada cümle grubuna yönelik olan bir ses kaydı yapılırsa başka herhangi bir sözcük ya da cümleyi tanıma olasılığı söz konusu olamaz.

Konuşma tanıma özelliklerinin kullanılabilmesi için sistemde bir ses kartı ve buna bağlı bir mikrofon bulunması gerekmektedir. Mikrofon ayarları tamamlandıktan sonra ses kaydetme işlemlerine geçilmiş ve bu işlem için KASİS 4.0 adında ses kaydetme programı kullanılmıştır.

Ses kaydetme işlemi için başka bir ses kaydedici yazılımı kullanmak da mümkün olmaktadır. Veri tabanında bulunan cümleler sırasıyla seslendirilmiş ve her 100 cümle bir konuşmacıya ait olacak şekilde çalışılmıştır (Şekil 2.1).



Şekil 2.1. Kasis 4.0 Programında Ses Kaydetme İşleminin Başlatılma Aşaması



Şekil 2.2 de Kaydedilen Bir Cümlenin Ses Sinyal Görüntüsü Görülmektedir.

Şekil 2.2. "Şu müzisyenler olağanüstü bir şekilde birbirine uyum sağlıyorlar" cümlesi Kullanılan mikrofonun düşük gürültü seviyesine sahip olması aranılan özelliklerindendir. Bu sistemde düşük gürültü seviyesine sahip SENNHEISER e8455 adlı dinamik bir mikrofon kullanılmıştır.

2.2. Öznitelik vektörlerinin eldesi

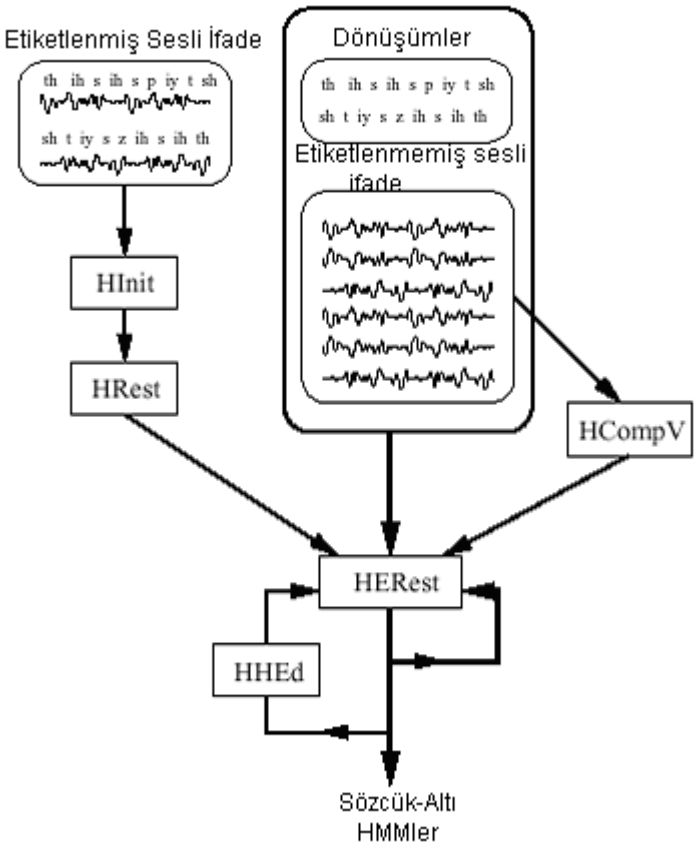
Eğitimde kullanılacak verilerin eğitim sırasında birden çok sayıda defa kullanılacağı göz önüne alınarak öznitelik vektörlerinin çıkarılması işlemi eğitim öncesinde yapılmıştır. Her ses dosyası için çerçevelere (frame) ait öznitelik vektörlerini içeren *.mfc dosyaları hazırlanmıştır. Bu işlem için HTK'nın HCopy aracı aşağıdaki parametrelerle uygulanmıştır:

- Çerçeve boyutu: 25ms
- Çerçeve ölçeklendirme fonksiyonu: Hamming
- Öznitelik vektörü: MFC
- Vektör ekleri: Birinci ve ikinci türev
- Cepstrum kanal sayısı: 12
- Cepstrum yükseltme: 22

2.3. Teklises eğitimi

Genel olarak geniş dağarcıklı bir sistem hedeflendiği için sözcük-altı akustik modeller kullanılmıştır (Şekil 2.3). Eğitim aşamasının ilk adımı teklises (monophone) modellerinin üretilmesidir. Teklises modelleri her ses (fonem) için bir adet olacak şekilde kullanılan tüm ses dosyalarının ortalama ve varyansları kullanılarak (HCompV programı ile) ilklendirilir. Oluşturulan SMM modelleri şu özellikleri barındırır:

- Durum sayısı: 5 (sadece 3 tanesi parametre içerir)
- Topoloji: Soldan sağa. Durum atlaması mevcut değil.
- Her durum için tek karışımlı Gauss olasılık dağılım fonksiyonu.



Şekil 2.3. Sözcük-Altı SMM'lerinin eğitimi

Ardından HHEd programı ile tüm modeller aynı anda eğitime tabi tutulur. Eğitim, HTK içinde gömülü yeniden tahmin (embedded reestimation) olarak adlandırılan bir yöntem ile gerçekleştirilir. Bu yöntem Baum-Welch parametre tahmin algoritmasının tüm eğitim verileri ve modeller için paralel olarak işletilmesini içerir. Gömülü yeniden tahmin her uygulandığında yeni bir akustik model kümesi elde edilir.

Teklises eğitimi bu şekilde dört kere yinelemeli olarak uygulanmıştır. Akustik model listesi 29 fonemden ve konuşma olmayan bölgeleri tanımlayan "sil" (silence) modelinden oluşmaktadır. Model adları karşılık geldikleri fonem ile küçük harf olacak şekilde gösterilmiştir. Ancak ASCII kümesinde olmayan Türkçe karakterler için en yakın ASCII karakterin başına 't' harfi eklenmiştir.

Örnek:

- ş → ts
- ç → tc

2.4. Üçlüseslere geçiş

Üçlüsesler bağlam bağımlı teklises modelleridir. Sol ve sağ yanındaki harfe göre her sesin birden çok sayıda akustik farklılık gösteren hallerinin ayrı akustik modellerle gösterilmesi amacıyla kullanılırlar. Tanıma sırasında da sözcük içindeki (veya sözcükler arasındaki) harf sıralamalarına göre uygun üçlüses modeli Viterbi algoritmasına sokulur. Bu yaklaşımın amacı akustik farklılıkların değerlendirilerek modellerin birbirinden ayrıklaştırılmasının sağlanmasıdır.

Üçlüses modelleri ait oldukları teklises modellerinden kopyalanarak elde edilir. Eğitim verileri içinde geçen tüm üçlüseslerin listesi, HLEd programı yardımıyla bulunur. Daha sonra bu listedeki her üçlüses için HHEd programı aracılığıyla ilgili teklisesin parametreleri kopyalanır.

Üçlüses modelleri elde edildikten sonra teklises modellerinde olduğu gibi HHEd programı ile gömülü yeniden tahmin işlemi yinelemeli olarak uygulanır. Sistemde kullanılan modeller için 3 adım yineleme gerçekleştirilmiştir.

2.5. Karar ağacı birleştirimi

Her fonem için sağ ve sol bağlamlarına bakarak farklı bir model oluşturmak, akustik açıdan anlamlı olsa da, istatistiksel olarak model başına düşen eğitim verisini azaltmaktadır (Jelinek, 1997:201). Özellikle nadir görülen üçlüsesler için yetersiz seviyede eğitim verisi bulunmaktadır. Bu nedenle Karar Ağacı Birleştirimi (KAB – DTC) algoritması kullanılmaktadır. Algoritmanın temeli üçlüseslerin ortak özelliklerinin kullanılarak bazı parametrelerinin paylaştırılması (yani ortak olarak eğitime tabi tutulması) prensibine dayanmaktadır.

```
QS "L_v_ince1"  {"i-*","e-*"}
QS "R_v_ince1"  {"*+i","*+e"}
QS "L_v_ince2"  {"tu-*","to-*"}
QS "R_v_ince2"  {"*+tu","*+to"}
QS "L_v_kalin1" {"ti-*","a-*"}
QS "R_v_kalin1" {"*+ti","*+a"}
QS "L_v_kalin2" {"u-*","o-*"}
QS "R_v_kalin2" {"*+u","*+o"}
QS "L_nasal"    {"n-*","m-*"}
QS "R_nasal"        {"*+n","*+m"}
QS "L_voicd_st" {"g-*","d-*","b-*"}
QS "R_voicd_st" {"*+g","*+d","*+b"}
QS "L_unvcd_st" {"k-*","t-*","tc-*","p-*"}
QS "R_unvcd_st" {"*+k", "*+t", "*+tc", "*+p"}
QS "L_fric_vcd" {"z-*","v-*"}
QS "R_fric_vcd" {"*+z","*+v"}
QS "L_frc_uvcd" {"s-*","ts-*","f-*"}
QS "R_frc_uvcd" {"*+s","*+ts","*+f"}
QS "L_whisper"  {"h-*"}
QS "R_whisper"          {"*+h"}
QS "L_oth_cons" {"y-*","r-*","l-*","c-*"}
QS "R_oth_cons" {"*+y","*+r","*+l","*+c"}
```

Yukarıda görüldüğü gibi sol veya sağ bağlamın ince ünlü, fısıltı, sürtünmeli ses olması gibi özellikler birleştirme kriteri olarak verilmiştir (Acar, 2001:68). Bu kriterlere göre tüm modellerin parametre içeren 2. 3. ve 4. durumları birleştirim işlemine sokulmuştur. HHEd programı içinde yer alan bu özellik sayesinde 26 bin olan tüm üçlüses modellerinin sayısı 1585'e indirilmiştir. Bu aşamadan sonra, yeni modellerin eğitimi için 3 yinelemeli gömülü yeniden tahmin eğitimi yapılmıştır.

2.6. Karışım sayısının artırılması

Akustik modeller ilk aşamadan itibaren tek Gauss karışımlı olarak kullanılmaktaydı. Ancak yeterli eğitim verisi olduğunda özellikle konuşmacı bağımsız akustik modellerin elde edilebilmesi için ses farklılıkları ve değişik söylenişlerin doğru bir şekilde modellere aktarılması amacıyla birden çok sayıda Gauss karışımı kullanılması gerekmektedir. Bu nedenle her üçlüses modeli için karışım sayısı HHEd programı kullanılarak 3'e çıkarılmıştır. Yeni oluşturulan modeller 3 yinelemeli eğitime tabi tutulmuştur.

2.7. Çalışma ortamı ve elde edilen dosyalar

HTK açık kaynak kodu ile dağıtılan bir araştırma kütüphanesidir. Windows platformu için uygulama programları olarak derlenmiş dosyalar web sayfasından indirilerek bilgisayara kurulmuştur. HTKBook (Hidden Markov Toolkit Book) adllı kitaptan faydalanılarak akustik model üretimi için yapılması gereken çalışmalar takip edilmiştir. İşlemler sırasında gerek olan metin dosyaları hazırlanmıştır. Bunlar ses dosyalarının listesi, metin içeriklerinin listesi, sözcüklerin hangi fonemlerden oluştuğunun listesi gibi bilgileri içeren dosyalardır. Bazı veri dosyalarının oluşturulması için Perl script dilinden faydalanılmıştır.

Akustik model üretimi bir perl dosyasının yukarıda anlatılan işlemleri sırayla yapması sayesinde gerçekleştirilmiştir. Elde edilen akustik modeller iki dosya olarak tanıma çekirdeğinin kullanımına sunulmuştur:

- hmmdefs.txt: SMM'lerin parametrelerinin tanımlandığı dosya.
- triphoneList.txt : SMM dosyalarının listesini tanımlayan dosya.

3. KONUŞMA TANIMA ÇEKİRDEĞİ VE ALGORİTMALARIN TANITIMI

Konuşma tanıma sırasında en çok kullanılan yöntem Viterbi algoritmasıdır. Viterbi algoritması genel anlamıyla bir dinamik eşleştirme (dynamic warping) yöntemidir.

3.1. Viterbi Algoritması

Tanıma için kullandığımız SMM modelleri birer sonlu durum makinesi (finite state machine) olarak görülebilir. Sonlu durum makineleri için en iyi olurluk değerini bulmaya yarayan bir optimizasyon algoritması olarak Viterbi algoritması konuşma tanıma sırasında en çok kullanılan yöntemdir (Becchetti and Ricotti, 1999:310).

Birkaç sözcükten oluşan bir dağarcığa sahip ve her sözcük için bir SMM'nin olduğu bir konuşma tanıma senaryosu düşünelim. Elimizde tanınacak konuşma verisinden elde edilen ve SMM'lerin eğitiminde kullanılan yöntemle elde edilmiş öznitelik vektörleri dizisi olsun. Bunlara gözlem vektörleri diyeceğiz.

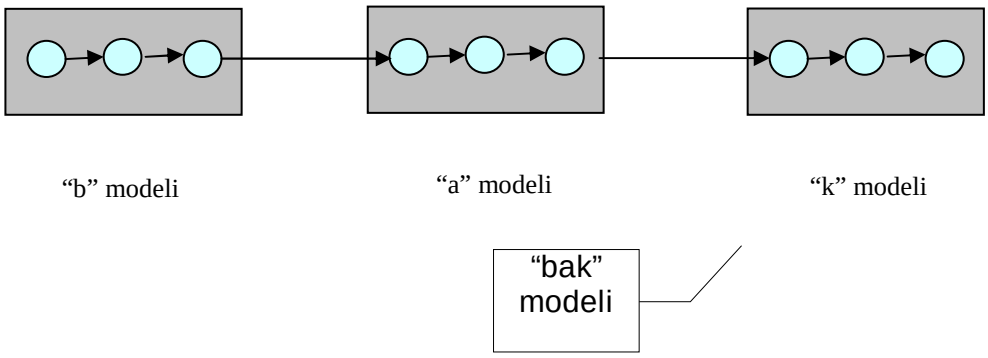
Algoritmayı şu şekilde tanımlayabiliriz:

1. Her sözcük modeli için:
2. Modelin ilk durumunu aktifleştir. Bu duruma ait skoru sıfırla.
3. Her gözlem vektörü için (sırayla)
4. Modeldeki tüm aktif durumlar için, eldeki gözlem vektörünün bu durum tarafından üretilme olurluğunun logaritmasını durum skoruna ekle.
5. Her durum için durum skorunu bağlantısı bulunan tüm durumlara ilet. (Skor iletilen durumlar otomatik olarak aktifleşir)
6. Her durum için kendisine iletilen skorlardan en yükseğini, bir dahaki gözlem vektöründe kullanılmak üzere seç.
7. Son duruma ait seçilen skor SMM'nin eldeki gözlem vektörleri için vereceği en iyi olurluk değerini tutmaktadır.
8. Tüm SMM'ler arasında en yüksek genel olurluk değeri veren modeli tanınan sözcük olarak tanı.

Skor hesaplamalarında logaritma kullanılmasının nedeni olasılık değerlerinin çarpılması gerekmesidir. Kayan nokta sayılar arasında çarpım yapmak yerine logaritmalarının toplanması daha basit bir işlemdir. Düşük olurluk değerlerinde sürekli olarak çarpım yapılması, ayrıca sayı hassasiyetinin azalmasına da neden olabilecektir.

Geniş dağarcıklı tanıma

Viterbi algoritmasının tanıtıldığı yukarıdaki örnek her sözcük için bir SMM modeli olduğunu varsaymaktaydı. Bu yöntem sözcük sayısı arttığında ya da sistem dağarcığına yeni bir sözcük eklenmeye çalışıldığında sorunlar yaratmaktadır. Bu nedenle sözcük modelleri yerine fonem modelleri kullanılmıştır. Fonem modeli(fonem tabanlı sesli ifade tanıma) kullanılarak herhangi bir sözcük için bütünleşik bir SMM oluşturulması, sözcüğün içerdiği fonemlere ait SMM'lerin sırayla eklenmesiyle sağlanmaktadır (Şekil 3.1).

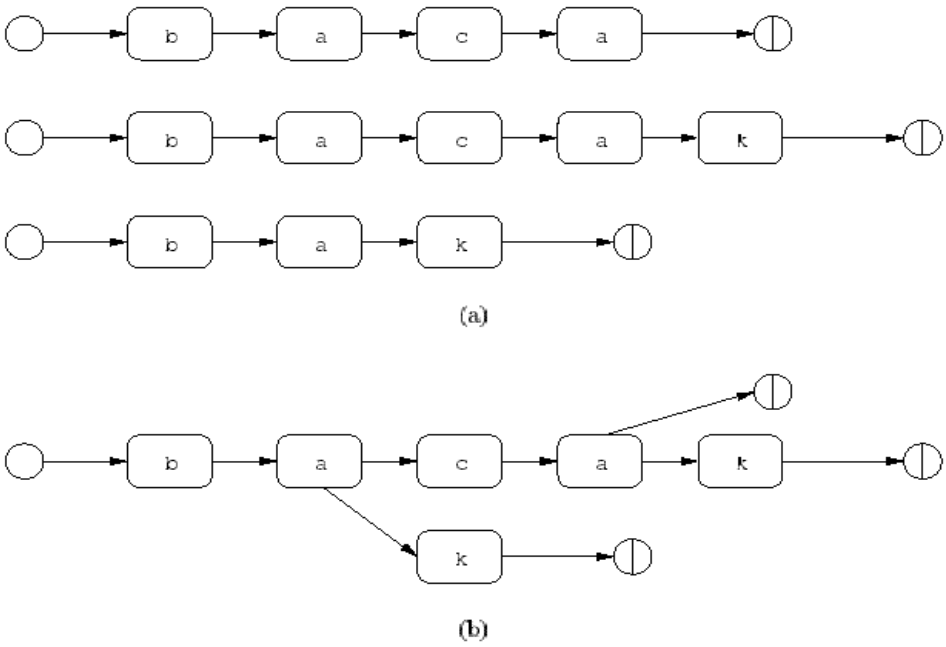


Şekil 3.1. SMM'leri ekleyerek oluşturulan model

Bütünleşik SMM ve Tanıma Ağı

Geniş dağarcıklı tanımada sözcük-altı akustik modellerin kullanımının yanısıra başka bir uygulama gereksinimi de tanınacak sözcük sayısının tanıma süresi üzerindeki olumsuz etkisinin giderilmesidir. Eğer dağarcıktaki her sözcük için ayrı ayrı bütünleşik SMM

oluşturulur ve hepsi için Viterbi algoritması uygulanırsa tanıma süresi sözcük sayısıyla (uzunluklarına göre) doğru orantılı olarak artacaktır (Young, 2002:16).



Şekil 3.2. a) Bütünleşik SMM Üzerinden Tanıma B) Sözcüklerin Ortak Harflerinden Faydalanma

Bu sorunun çözümü hiyerarşik olarak verimli bir yapıda ortak fonemlerin kullanılmasını sağlayan ağaç yapısıdır. Şekil 3.2 de ağaç yapısındaki bütünleşik SMM'nin ifade avantajı görülebilir: Şekilde baca, bacak ve bak sözcükleri için alternatif bütünleşik SMM'ler görülmektedir. (a) şeklindeki bütünleşik SMM üzerinden tanıma, yukarıda belirtildiği gibi verimsizdir. (b) şeklinde gösterilen ifade biçimi ise sözcüklerin ortak harflerinden faydalanmaktadır. Bu nedenle hem ifade daha verimli olmuştur, hem de bu bütünleşik SMM ağı üzerinden tanıma yapılırken Viterbi algoritması her gözlem vektörü için "b" modelinin durumları üzerinden sadece bir kez geçecektir. İlk şekildeki SMM için bu sayı 3'tür.

Düğümler

Şekil 3.2 (b)'de "a" ve "k" etiketi barındıran ikişer tane düğüm (node) bulunmaktadır (bkz. Şekil 3.2). Kullandığımız ağaç yapısındaki düğümler şu aşamadan itibaren birebir SMM'ler olmayacak, tanıma ağı düğümleri olarak adlandırılacaktır. Her tanıma ağı düğümü aşağıda gösterilen bilgileri taşır:

- Düğümün bir sözcük sonu olup olmadığı, öyle ise sözcüğün ne olduğu.
- Düğüme ait harf / fonem (Üçlüses adı: örnek: b-a+k)
- Düğüme ait SMM için bağlantı.
- Düğüme ait olurluk skoru.

- Varsa düğümün eş (peer) düğüm bağlantısı.
- Varsa düğümün çocuk (child) düğüm bağlantısı.

Tanıma Ağacının Oluşturulması

Tanıma ağacının oluşturulması sistem dağarcığındaki tüm sözcüklerin yukarıda anlatılan ağaç yapısı içinde yer alabilmesi için gerekli düğümlerin oluşturulmasıdır. Bu işlemin işleyişi aşağıdaki gibidir:

1. Başlangıç ve bitiş düğümlerini oluştur.
    9. Bu düğümleri sessizlik SMM'siyle bağlantılandır.
    10. Dağarcıktaki her sözcük / cümle (W) için:
    11. Üçlüses açılımını bul (örnek: “bak” sözcüğü için sırasıyla b+a, b-a+k, a-k üçlüsesleri kullanılır)
    12. İlkini seç. üçlüses = ilk üçlüses
    13. Tarama için başlangıç düğümünden başla. düğüm = BaşlangıçDüğümü 7. düğüm.Çocuğuvar (üçlüses) doğru olduğu sürece
    14. düğüm = düğüm.ÇocukDüğüm
    15. üçlüses = bir sonraki üçlüses.
    16. Geriye kalan ve ağaçta bulunmayan üçlüsesler için
    17. düğüm.ÇocukDüğümEkle (üçlüses)
    18. düğüm = düğüm.ÇocukDüğüm
    19. düğüm.smm = üçlüsese ait SMM
    20. üçlüses = bir sonraki üçlüses.
    21. düğüm.Sözcük = W

Bu işlem tamamlandığında tanıma algoritmasının tanıma yapabileceği ağaç yapısındaki tanıma ağı hazır olmuş olacaktır. Bu işlemin yapılmasından önce sistemde kullanılacak olan SMM'lerin yüklenmiş ve hazır olması gerekmektedir. Her düğüm, temsil ettiği üçlüsese ait modele bağlantılandırılır. HTK ile yapılan model eğitimi çalışmalarında HHEd programı ile, olası tüm üçlüsesler için akustik model üretilmesi sağlanmıştır. Böylece tanıma ağı, dağarcığa eklenebilecek herhangi bir sözcük veya cümleyi oluşturabilecek şekilde çalışabilir.

Tanıma İşleminin Gerçekleştirilmesi

Tanıma algoritması daha önce anlatılan Viterbi algoritmasının ağaç yapısındaki bütünleşik SMM yapısına uyarlanmış halidir. Viterbi algoritmasında bahsi geçen durum skorları her durum için marka (token) adı verilen yapılarda saklanmaktadır. Markalar algoritmanın kolay işlemesi açısından ikili diziler şeklinde saklanır ve ardışık gözlem vektörleri için sırayla kullanılır (her durum için iki adet marka bulunur, tanıma sırasında bunlardan bir tanesi o andaki skoru, diğeri bir sonraki gözlem için seçilen skoru saklar).

Kullanılan tanıma yönteminin bir diğer özelliği de tanıma süresinin kısıtlanması için ağaç budama tekniği kullanılmasıdır. Bu teknik tanıma sırasında belirli bir gözlem vektörü için, o an bulunan en yüksek düğüm skorundan budama eşiğine göre daha düşük skor veren düğümlerin tanıma işleminden çıkarılması esasına dayanır. Tanıma oranını düşürmeden, yeterince uygun olmadığı görülen sözcük hipotezlerinin elenmesini sağlar.

Algoritma aşağıda anlatılan şekilde işler:
1. AktifDüğümListesini sıfırla
2. AktifDüğümListesine başlangıç düğümünü ekle.
3. Ağaç düğümlerinin markalarını sıfırla.
4. Sırayla tüm gözlem vektörleri (V) için:
5. Her düğüm = elemanlar (AktifDüğümListesi) için:
6. düğüme ait her durum (s) için (1..3):
7. Olurluğu hesapla. olurluk = P(V | s). Gauss karışımlarından elde edilir.
8. s.marka.olurluk += olurluk.
9. Elde edilen markayı bir sonraki duruma (s+1) aktar.
10. Aktarılan marka durumun (s+1) alternatif markasının skorundan iyiyse kabul edilir.
11. Son durum ise (s = 3), bulunulan düğümün tüm çocuk düğümleri (çocuk) için:
12. çocuk düğümü AktifDüğümListesine ekle.
13. Durumun markasını düğümün ilk durumuna aktar (Kabul işlemi 10. maddedeki gibidir).
14. düğüm sözcük sonu ise düğümün son durum markasını bitiş düğümüne (boşluk modeli içeren) ilet. Kabul işlemi madde 10 daki gibidir.
15. Şu ana kadar rastlanan en yüksek olurluk değerini bul. eniyiskor = maksimum (düğüm skoru).
16. Düğüm durumları için marka kümelerini yer değiştir.
17. AktifDüğümListesindeki tüm düğümler için:
18. Düğüm skoru eniyiskordan eşikdeğeri veya daha fazla düşük ise düğümü Aktif Düğüm Listesinden çıkar.

Tanıma sırasında ağaç budaması eşik değeri olarak 250 kullanılmıştır. Bu algoritma tamamlandığında bitiş düğümündeki son duruma ait marka tanınan sözcük (veya cümlenin) bilgisini saklıyor olacaktır.

4. KTM v(1.1)'NİN KONUŞMA TANIMA ÖZELLİĞİ

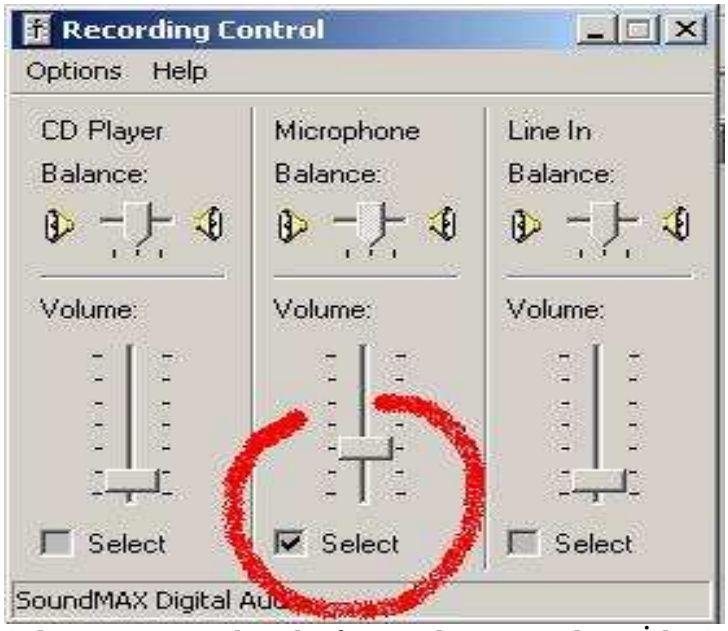
Konuşma tanıma için gerekli olan bütün çalışma ve algoritmaların tamamlanmasıyla beraber yine hazırlanan metin editörü içerisine bu teknoloji entegre edilerek hazırlanan KTM (Konuşma Tanıyan Metin editörü) yazılımının kullanıcı arayüzü tamamlanmış olur.

KTM'nin en önemli özelliği konuşma tanıma sayesinde metin yazılması ve komut verilmesi işlemlerinde kullanıcıya sesiyle giriş yapma imkanı sağlamasıdır. Konuşma tanıma sistemiyle ilgili bilgiler aşağıdaki başlıklar altında toplanmıştır.

- Mikrofon Ayarları
- Fiş Tanıma
  - Fiş Listesi Düzenleme
- Komut Tanıma

4.1. Mikrofon Ayarları

Konuşma tanıma özelliklerinin kullanılabilmesi için sistemde bir ses kartı ve buna bağlı bir mikrofon bulunması gerekmektedir. Tercihen mikrofonun düşük gürültü seviyesine sahip olması iyi olacaktır. Sistem ses ayarlarından kayıt sırasında seçili kaynağın mikrofon olması sağlanmalıdır (Şekil 4.1). Ses mikrofon girişi ses yüksekliği de uygun şekilde ayarlanmalıdır. Deneme olarak ses kaydedici ile kayıt yapılarak konuşmanın rahat anlaşılabilir olduğundan emin olunmalıdır.



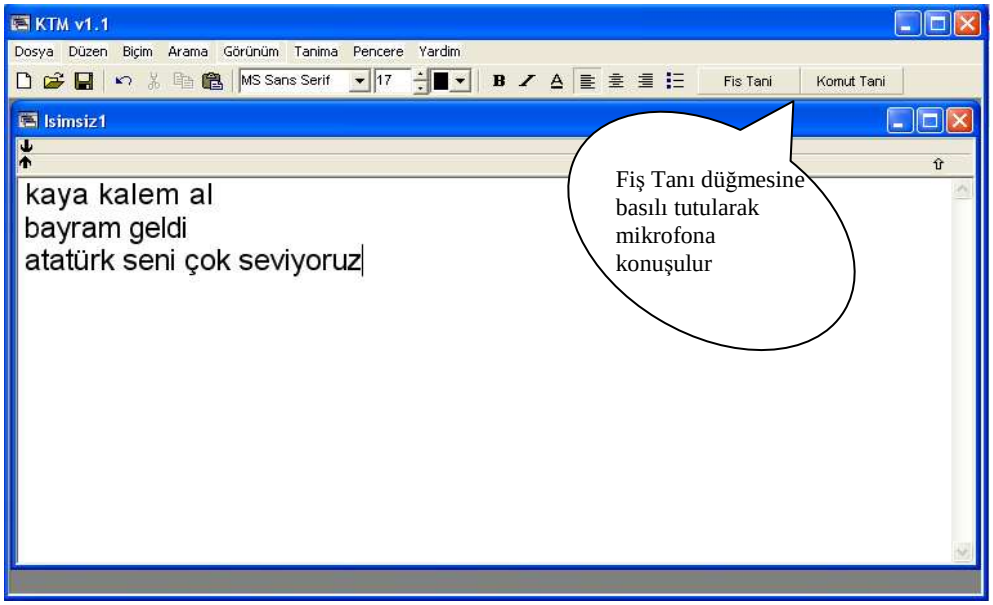
Şekil 4.1. Windows 98 Ortamında Mikrofon Ayarlarının Yapılması İşlemi

4.2. Fiş tanıma

Konuşma tanıma metin editöründe konuşma tanıma ifadesi yerine fiş tanıma ifadesinin kullanılma sebebi, KTM'de ilköğretim birinci sınıf öğrencilerine ilkokuma yazma işlemini öğretmek olduğu için, ilköğretimde daha önce kullanılan fiş cümlelerine karşılık gelmesinden kaynaklanmaktadır. Buradaki fiş kelimesi bilgisayar tarafından tanınması istenilen herhangi bir metne karşılık gelmektedir.

Tanıma çubuğundaki Fiş Tanı düğmesi ile fiş tanıması yapılarak, tanınan fişin (sözcükler/cümleler) seçili pencere içine yazdırılması sağlanmaktadır (Şekil 4.2). Bunun için aşağıdaki adımlar gerçekleştirilmelidir.

1. Fiş Tanı düğmesine tıklamak (farenin sol düğmesini bırakmamak)
2. Mikrofona yazdırmak istenilen fişi söylemek.
3. Fare düğmesini bırakarak tanıma işlemini başlatmak.

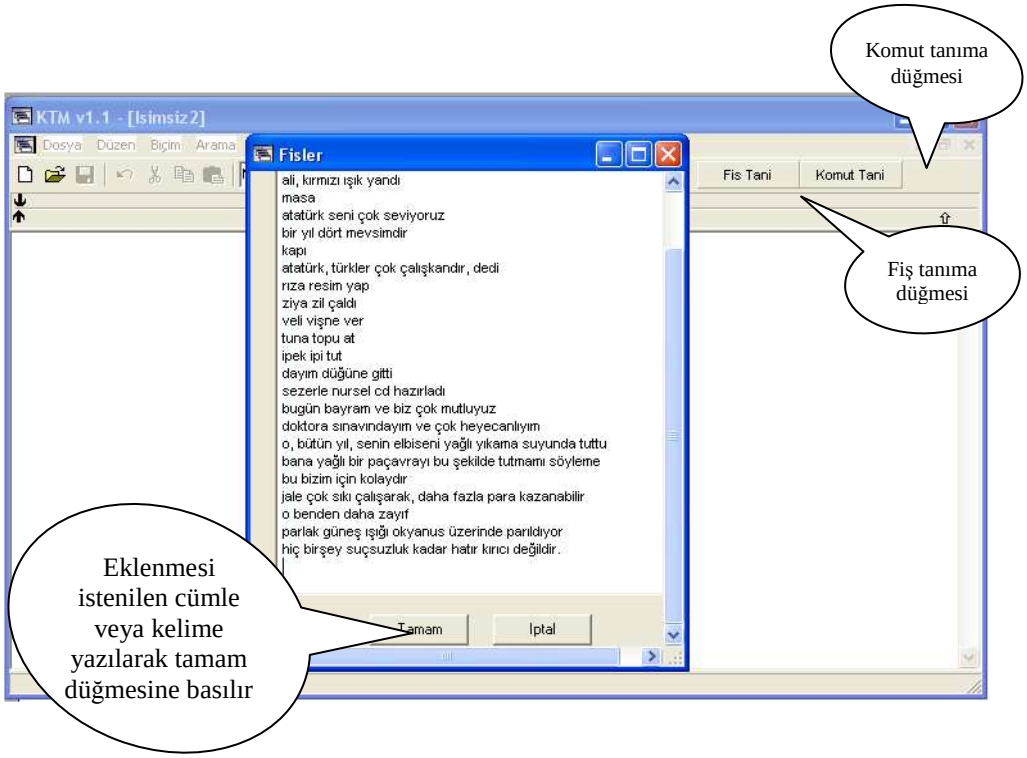


Şekil 4.2. Fiş Tanıma Düğmesiyle Ekrana Yazdırılmak İstenilen Cümlenin Seslendirilme Aşaması

Kısa bir süre içinde söylenmiş olan fiş tanınarak metin penceresi içine eklenecektir. Dağarcıktaki fişleri görmek ve düzenlemek için Fiş Listesi penceresi kullanılmaktadır.

Fiş Listesi Düzenleme

Fiş listesi penceresini açmak için tanıma menüsünden Fiş Listesi seçeneği seçilmelidir. Pencerede dağarcıktaki fişler liste halinde görüntülenebilmektedir. Liste üzerinde herhangi bir fiş seçilerek istenen değişiklikler yapılabilmektedir. Bu liste üzerinde her satır, bir fişe karşılık gelmektedir (Şekil 4.3.).



Şekil 4.3. Fiş Listesine Her Hangi Bir Cümle Ekleme Veya Değiştirme İşlemi Aşamaları

Yapılan değişiklikleri onaylayarak fiş listesini kaydetmek için Tamam, kaydetmeden çıkmak için İptal düğmesi tıklanılır.

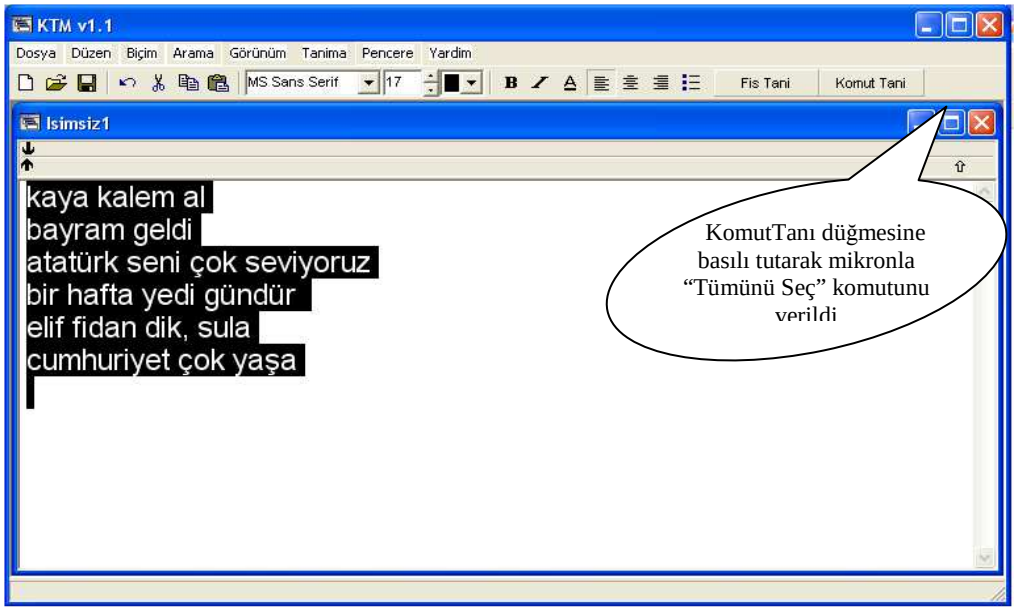
4.3. Komut tanıma

Tanıma araç çubuğu üzerindeki Komut Tanı düğmesine tıklanarak menü komutları uygulanabilmektedir. Bunun için aşağıdaki adımlar gerçekleştirilmelidir.

1. Komut Tanı düğmesine tıklamak (farenin sol düğmesi bırakılmadan).
2. Mikrofona uygulamak istenilen komutu söylemek.
3. Fare düğmesini bırakarak tanıma işlemini başlatmak.

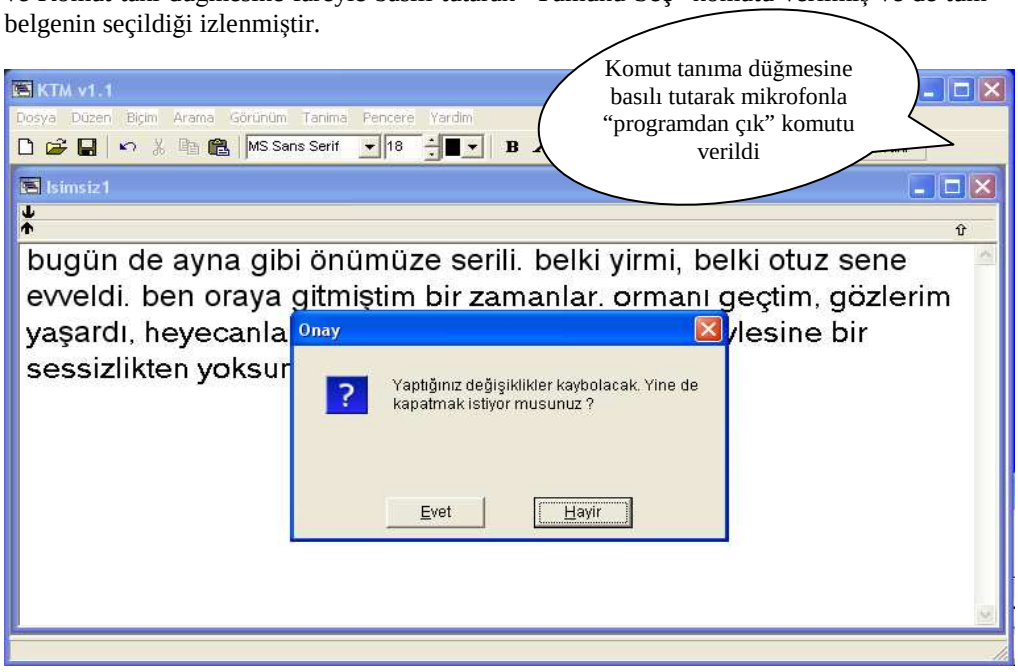
Şekil 4.4.de “tümünü seç” komutunun, Şekil 4.5. de ise “programdan çık” komutunun çalışma düzenleri görülmektedir.

KTM menüsündeki tüm komutlar bu şekilde uygulanabilmektedir. Tanınmış olan komut durum çubuğunda birkaç saniye görüntülenecektir.



Şekil 4.4. Komut Tanı Düğmesine Mikrofonla "Tümünü Seç" Komutunun Verilmesi

Şekilde de görüldüğü gibi metin editörüne mikrofonla seslendirilen bu cümleler okunulmuş ve Komut tanı düğmesine fareyle basılı tutarak "Tümünü Seç" komutu verilmiş ve de tüm belgenin seçildiği izlenmiştir.



Şekil 4.5. Mikrofonla "Programdan Çık" Komutunun Verilmesi

Aynı şekilde menüler de bulunan tüm komutların (yaklaşık 30) bu şekilde uygulanması mümkündür olmaktadır. Fiş Listesinin alabileceği satır sayısı 10.000 kadardır. Yani 10.000 ayrı cümlenin bu sistemde tanınması mümkündür.

5. SONUÇLAR

Sesli metinlerin işlenmesi doğal dil, sesbilim, akustik ve sayısal sinyal işleme gibi alanlarla ilgilidir. Dilbilim teorileri dilin fonemlerini ve onun anlam bütünlüğünü; akustik teoriler, fonemlerin özelliklerini; sinyal işleme ise ses sinyallerinin yapısını incelemektedir. Türkiye'de konuşma tanıma alanında yapılan çalışmalar çok fazla değildir. İnsanlığın hizmetine sunulabilecek alanlardan olan bu sistem insan-makine etkileşiminin en güzel örneklerinden biridir. Özellikle konuşmacıdan bağımsız olarak çalışabilen bağlı konuşma tanımayla beraber geniş dağarcıklı konuşma tanıma sistemlerini oluşturabilme bu alanda yapılabilecek en iyi çalışmalardan olacaktır. Konuşma tanıma hemen hemen her alana uygulanabilir. Ancak bu teknolojiyi hazırlamak zor olduğundan ve uzun zaman gerektirdiğinden çok fazla tercih edilmemektedir. Oysa bilgisayara konuşarak bir şeyler yazdırmak, konuşarak bir şeyleri yaptırmak oldukça faydalı bir çalışmadır. Özellikle engelliler (görme, duyma ve bedensel) için bu tarz çalışmalara ağırlık verilmelidir. Konuşma tanıma teknolojisiyle hazırlanan çalışmalar engelliler için yeni bir özgürlük alanı olacaktır.

Konuşma Tanıma sistemleri dilden bağımsız olarak hazırlanabilmelidir. Bu amaç doğrultusunda yeni algoritmalar geliştirilmelidir. Ülkemizde konuşma tanıma çalışmalarında yapay zeka tekniklerinden yapay sinir ağlarına ait çeşitli algoritmalar üzerinde çalışılmıştır. SMM kadar çok sıklıkla kullanılmasa da yapay sinir ağlarıyla ilgili yapılan çalışmalarda, ses eğitme aşamasının oldukça uzun sürdüğü yapılan araştırmalarda görülmüştür. Kelime veya cümle sayısının artmasıyla bu başarı oranı daha da düşmektedir. Diğer ülkelere bakıldığında konuşma tanıma sitemlerinde genel olarak SMM yapısı kullanılmıştır. Bunun yanı sıra diğer yapay zeka teknikleriyle de çalışılmıştır (Genetik algoritma, fuzzy logic, uzman sistemler). Öyle ki hybrid sistemler denilen karma veya melez sistemlerle bu çalışmalara ağırlık verilebilir. Hem yapay sinir ağı ve hem SMM sistemleri birlikte kullanılabilir. Bu örnek diğer yapay zeka teknikleriyle de çoğaltılabilir. Ülkemizde konuşma tanıma teknolojilerinde yapay sinir ağları haricinde diğer yapay zeka teknikleriyle çalışılmadığı gözlenmiştir. Ayrıca konuşma tanıma çalışmaları için hazır ses veri tabanı kütüphanesi oluşturulmalıdır. Tubitak bu konuyla ilgili olarak çalışmalarına devam etmektedir. Türkçe konuşma tanıma yapabilen sistemlerin geliştirilmesi üzerine daha çok araştırma yapılmalıdır. Özellikle konuşmacıdan bağımsız bağlı konuşma tanımayı sağlayan çalışmalar daha kullanışlı olacaktır. Örneğin mahkeme duruşmalarında, emniyet sorgularında, zabıt işlemlerinde hep karşılıklı konuşmaların anında bilgisayara yazılması söz konudur. Böyle bir alana yönelik yapılan çalışma oldukça kullanışlı olacaktır. İnsanlığı daha rahat bir çalışma ortamına kavuşturabilmek için konuşma tanımayla ilgili çalışılabilecek alanların çeşitliliğinin artırılması gerekmektedir.

Hazırlanan KTM v.1.1. yazılımında fiş listesine ilk olarak 60 cümle eklenmişti. Amaç sadece ilköğretim 1. sınıf öğrencilerine yönelik öğretilen fişlere ait cümlelerin program tarafından tanınacak cümleler olmasıydı. Bu nedenden dolayıdır ki öncelikle tanınması istenen ilk 60 kelime veya cümlelerin içeriği ilköğretim 1. sınıfta öğrencilere öğretilen fişlerin listesiydi. Yine bu sebepten yazılım sözlüğü, yazılım içerisinde fiş listesi adlı alt menü olarak kullanılmıştır. Ancak KTM içindeki fiş listesinin değiştirilebilir-düzenlenebilir

özellik taşımasından dolayı amacın oldukça dışına da çıkmak mümkün olmaktadır. Aslında fiş listesi yazılımın sözlüğünü oluşturmaktadır. Fiş listesine 10.000 farklı cümle eklenebilmektedir. Bu da 10.000 farklı cümlenin KTM tarafından tanınıyor olması anlamına gelmektedir. Bu rakam sistem programlanırken sınır 10.000 alındığından dolayı 10.000 farklı cümleden bahsedilmektedir. Bu sınırın daha da yukarıya çekilmesi mümkündür. Konuşmacıdan bağımsız sürekli konuşma tanımayı sağlayan KTM v1.1, EK-2 de de gösterilen komutları sesli olarak tanımaktadır. Fiş listesindeki cümle sayısını artırarak KTM v1.1 yazılımının, konuşma tanımadaki başarısının nasıl olacağı merak edilmiştir. Bu maksatla programdaki fiş listesine, 60 cümle yanında 100 cümle daha eklenmiştir (bu cümleler "GÖÇEBE" adlı romandan alınmıştır). KTM'ye mikrofonla önce yeni 100 cümle okunmuş, sonra 160 cümle okunmuş ve yazılımın konuşma tanıma başarısı 1. etapta test edilmiştir. Bu test etme aşamasında ve bu yazılımla çalışma aşamasında gürültü seviyesinin çok fazla olmaması gerekmektedir. Ayrıca herhangi bir gürültünün olmaması daha da iyi bir sonuç verecektir. Test sonuçları konuşma tanımada başarı sağlandığını göstermiştir. Fiş listesinde 160 cümleyle yazılımı test etmenin yetersiz olacağı düşünülerek 2. etap olarak 100 cümle daha yazılımın fiş listesine eklenmiştir. Hem yeni oluşturulan cümlelerle hem de var olan önceki cümlelerle (toplam 260 cümle) KTM v1.1 yazılımının konuşma tanıma oranı tekrar test edilmiş ve yine tanıma başarısının oldukça iyi sonuçlar verdiği gözlemlenmiştir. Farklı konuşmacılarla yeniden test edilen yazılım başarısı yine aynı oranda başarı göstermiştir. Yine cümle sayısının yeterli olmadığı düşünülerek farklı 100 cümle daha yazılımın fiş listesine eklenerek (toplam 360 cümle) yine 3. etap olarak yazılım başarısı test edilmeye çalışılmıştır. Gözlem sonuçları yine aynı çıkmış ve yazılımın başarı oranı değişmemiştir. 4. etapta 100 farklı cümle daha yazılımın fiş listesine eklenerek yeniden yazılımın konuşma tanıma başarısı test edilmeye çalışılmıştır. Elde edilen sonuç öncekilerden farklı değildir. Yani 4. etapta da (toplam 460 cümle) KTM v1.1 konuşma tanıma başarısı oldukça yüksektir. Yine bu eklenen yeni cümleler farklı konuşmacılarla tekrar test edilmiş ve yine aynı başarı oranı gözlenmiştir. Yapılan testlerde % 100'e yakın başarı gözlenmiştir Oluşan bazı hatalar ise fiş tanıma düğmesine zamanında basılmadığı veya zamanından önce fiş tanıma düğmesinden uzaklaşıldığı veya mikrofona duyarlılıkla beraber ortamın gürültüsü olmaktadır.

Türkçe olarak hazırlanmış konuşma tanıma uygulamaları yok denecek kadar azdır. Teknolojik bir devrim olabilecek nitelikteki bu çalışmalara eğitim alanında sıklıkla yer verilmelidir. İlköğretim birinci sınıf öğrencilerine ilkokuma yazma öğretimine yönelik hazırlanmış olan KTM uygulaması okul öncesi eğitimde de tercih edilebilir niteliktedir. Bu çalışmayla öğrenci derse daha çabuk konsantre olacaktır. Çünkü bilgisayar teknolojisiyle daha eğitiminin ilk aşamalarında tanışmış olan öğrenci kendini daha mutlu hissedecek ve derse katılımı daha fazla olacaktır. Kısacası hazırlanan yazılımın ilkokuma yazma öğrenme sürecinde öğrencilerin bireysel, zihinsel ve sosyal gelişimlerine katkı sağlayacağı düşünülmektedir.

KAYNAKLAR

Nabiyev, V. "Yapay Zeka", ISBN 975 347 985 9, ***Seçkin Yayıncılık San. Ve Tic. A. Ş.***, Ankara, 2005.

Öncül, N., "Kısa Eğitme Süreli Bir Konuşmacı Tanıma Dizgesi Tasarımı ve Gerçekleştirilmesi", Yüksek Mühendislik Tezi, ***Hacettepe Üniversitesi Fen Bilimleri Enstitüsü***, Ankara, (1993).

Gökhan, A., "Yapay Sinir Ağları İle Ayrık Türkçe Sözcüklerin Tanınması", Yüksek Lisans Tezi, ***Fırat Üniversitesi Fen Bilimleri Enstitüsü***, Elazığ, (1997).

Jelinek, F., "Statistical Methods for Speech Recognition", ISBN 0-262-1006-5, ***The MIT Preess Cambridge***, Massachusetts London, England, (1997).

Acar, D., "Triphone Based Turkish Word Spotting System", Master Thessis, ***The Department of Electrical And Electronics Engineering, The Middle East Technical University***, Ankara, (2001).

Becchetti, C., and Ricotti L. P., "Speech Recognition Theory and C++ Implementation", ISBN 0-471-97730-6, ***John Wiley & Sons Ltd,*** England, (1999).

Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., and etc. , "The HTK Book (for HTK Version 3.1)", ***Copyright (1995-1999) Microsoft Corparation, Copyright (2001-2002), Cambridge University Engineering Department***, Cambridge, (2002).