Türkçe Metin Seslendirme

Tuncay Şentürk
İTÜ Bilgisayar Mühendisliği Bölümü

Eşref Adalı
İTÜ Bilgisayar Mühendisliği Bölümü

{tuncay.senturk@gmail.com, adali@itu.edu.tr}

Özetçe

Bu çalışmada temel amaç, Türkçe metinlerin insan sesine dönüştürülebilmesi ve "Türkçe Metin Seslendirme" sisteminin geliştirilmesidir. Bu sistem geliştirilirken üç farklı yöntem incelenmiş, uygulanmış ve aralarındaki anlaşılırlık istatistiksel olarak ölçülmüştür. İlk olarak, "çift-ses (diphone) eklemeli yöntem" uygulanmıştır. Anlaşılırlığı düşük olmasa da doğallıktan uzak sonuçlar elde edilmiştir. Bunun üzerine, donanım maliyetinin de azalması ile, çift-ses eklemeye nazaran günümüz koşullarında daha kabul görmüş "hece eklemeli yöntem" geliştirilmiştir. Anlaşılırlık olarak ve ses kalitesinde olumlu yönde fark olduğu istatistiksel olarak ispatlanmıştır. Son olarak, ses süre ve şiddetinin değiştirilmesi suretiyle, vurgu ve tonlamada da başarılı sonuçlar elde edilmiştir. Tüm yöntemlerin ayrı ayrı anlaşılırlığının tespit edilebilmesi ve karşılaştırılabilmesi için; belirlenmiş cümleler, farklı yaş gruplarındaki insanlara dinletilmiş ve alınan cevaplara göre belirli formül yardımı ile yüz üzerinden puan verilecek şekilde hesaplama yapılarak, bir matriste sunulmuştur.

Bu çalışmada farklı Türkçe ses sentezleme yöntemleri karşılaştırılmış ve kullanıcı deneyleri ile kalite analizi gerçekleştirilmiştir. Ses sentezleme yöntemlerinin karşılaştırmalı incelenmesi ve yapılarla oynanmasına müsaade edilen bir biçimde sunulması (XML), bu makalenin önemli katkı sağlamış olduğu noktalardır.

Tüm çalışmalar için gerekli ses dosyalarının hazırlanması amacıyla önce Türk Dil Kurumunun ses veri tabanı kullanılmıştır. Daha sonra, yazılan program vasıtası ile MBROLA kütüphanelerinin kullanılması ile, tüm ses dosyalarının otomatik olarak oluşturulabilmesi sağlanmıştır. Oluşturulan bu ses dosyalarına, genlik dengeleme algoritması uygulanmış, ses dosyaları arasındaki en fazla ve en az genlik seviye farklılıkları aza indirgenerek anlaşılırlık arttırılmıştır.

Hazırlanan programın gevşek bağlaşımlı bileşenlerden (metinden XML geçişi ve XML'den ses oluşturulması) oluşabilmesi sağlanmış ve bu bileşenler kullanılarak kullanıcı arayüzü hazırlanmıştır.

Son olarak, görme engellilerin de ekran görüntüsü gerektirmeden kullanabileceği metin düzenleme program hazırlanmıştır.

Abstract

Turkish Text to Speech Synthesizer

The main purpose of this study is development of a "Turkish Text Synthesizer System which converts text, written in Turkish, to human voice. Three different methods are examined for developing this system, these three methods are implemented and their clarity is measured statistically.
First, the diphone concatenation method was applied. While the words were understandable, results were far from natural. Thus, considering the reduction of hardware costs in todays conditions the more accepted "syllable concatenation method" was developed. It is statisticaly proven that there is positive improvement with clarity and sound quality with this method. Finally, by changing the amplitude and duration of the sounds, more successful results were obtained for intonation. In order to determine and compare clarity of all methods set sentences were listened by different age groups and their answers were formulated to a score from 0 to 100, and the results were given in a matrix.

The Turkish Language Association's (TDK) database is used to prepare the necessary audio files in the begining of this study. Then, by means of a software program developed, MBROLA library was used to automatically create all the sound files. The amplitude balancing algorithm has been applied to these audio files, and clarity was increased by normalizing the maximum and minimum amplitude differences between sound files.

It is provided that, the system has loosely coupled components (text to XML and XML to speech), and using these components a graphical user interface is developed.

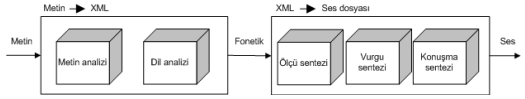
Finally, a text editing software program is developed to help the visually impaired edit text without the need for a screen image.

1. Giriş

Konuşma, insan haberleşmesinde en etkin yollardan birisidir. Teknoloji ilerledikçe makina-insan etkileşimi de önem kazanmış ve çeşitli yöntemler sunulmaya başlanmıştır. Metin seslendirme de bu yöntemlerden birisidir ve kullanıcının sürekli bilgi kaynağını izlemesi zorunluluğunu ortadan kaldırır.

Türkçe Metin Seslendirme Sistemi, Şekil-1'de gösterildiği gibi gevşek bağlaşımlı iki temel bileşenden oluşur. Aynı zamanda her iki bileşen de

farklı uygulamalarda bağımsız şekilde kullanılabilecek şekilde tasarlanmıştır. İlk bileşen metnin, dilbilimsel kurallar çerçevesinde, ses sinyallerine dönüştürülmek üzere belirlenecek bir biçime dönüştürülmesini sağlamaktadır. Bu çalışmada, bunun için XML kullanılmıştır. İkinci bileşen, birinci bileşence veya dış programlar aracılığıyla hazırlanmış XML katarının veya dosyasının ses sinyallerine dönüştürülmesini sağlamaktadır. Bu bileşen XML içeriğini tarayarak, gerekli ses dosyalarını birleştirme yolu ile en anlaşılır ses dosyasını üretmektedir.



Şekil 1: Metinden ses elde etme bileşenleri

Her iki bileşenin birbirinden oldukça basit bir biçimde ayrılması XML yapısı ile sağlanmıştır. XML, ses sentezleme yöntemlerinin karşılaştırmalı incelenmesini ve yapılarla oynanarak müsaade edilen şekilde sunulmasını sağlayarak, makalenin önemli katkı sağlamasına sebep olmuştur. Çalışmada ayrıca, kullanıcının bu iki bileşenin detayını bilmesini gerektirmeden, sadece girdiği metni seslendirebilmesi amacıyla, basit bir arayüz de hazırlanmıştır.

Tüm bu çalışmaların da kullanıldığı bir arayüz ile; görme engelliler için metin düzenleme programı da hazırlanmıştır. Bu program ile bilgisayar ekranına bakmadan, sadece tuştakımı ile komutların ve yazılan metnin seslendirilmesi sağlanabilmiştir.

2. Kaynak Taraması

Ses sentezleme açısından önemli sayılan çalışmalar, Çizelge-1'de dahil oldukları yöntemlerle birlikte gösterilmiştir.

MITalk, biçimlendirici (formant) temelli olup, günümüzde kullanılan birçok çalışmaya temel teşkil etmiştir[2, 3].

Çizelge-1: Önemli Metinden Ses Üretme Sistemleri

Çalışma	Yöntem	Tarih
MITalk	Biçimlendirici (formant)	1979
Infovox	Biçimlendirici (formant)	1982
Bell Labs TTS	Çift-ses, üçlü ses ekleme	1973
ETI Eloquence	Eklemeli	1988
CNET PSOLA	Çift-ses ekleme	1980'li yılların ortaları
Festival TTS	Çift-ses ekleme	1990'lı yılların sonları
MBROLA	Çift-ses ekleme	1990'lı yılların sonları
Whistler		
GVZ	Hece ekleme	2000'li yıllar

1982 yılında, İsveç Royal Institute of Technology'de çok dil destekli (multilingual) olarak geliştirilmiş ticari bir uygulama olan Infovox metinden ses üretme anlamında en önemli projelerden biridir. İlk sürümlerinde basamaklı biçimlendirici (cascade formant) yöntemi kullanılmaktaydı ve İngilizce metin seslendirme aşamasında üretilen seslerde İsveç aksanı ön plandaydı. Daha sonra çıkarılan sürümlerinde ise çift-ses ekleme (diphone concatenative) yöntemi kullanılmıştır[4, 5].

Bell Labs TTS çift-ses (diphone) ve üçlü ses (triphone) ekleme (concatenative) yöntemine dayanmaktadır ve İspanyolca, İtalyanca, Rusça, Romence, Çince ve Japonca desteği bulunmaktadır [6].

SoftVoice, TTS konusunda 25 yıldan fazla tecrübesi olan SoftVoice firması tarafından geliştirilmiş ve SAM (Software Automatic Mouth) olarak bilinmektedir. Genellikle Commodore C64, Amiga ve Atari bilgisayarlarında çoğul ortam ürünü olarak kullanılmıştır ve 1980'li yılların başlarında kişisel bilgisayarlar için tercih edilen ilk ticari TTS uygulamalarından birisi olmuştur [7].

ETI Eloquence, Eloquent Technology (ABD) tarafından geliştirilmiş, eklemeli yöntem kullanan, çoklu dil desteği sunan bir sistemdir [8]

Festival, Edinburgh Üniversitesi Ses Teknolojileri Araştırma Merkezi'nde Alan Black ve Paul Taylor

tarafından 90'lı yılların sonlarında geliştirilmiştir. İkili ses ekleme yönteminin uygulandığı sistem dilden ve platformdan bağımsız çalışmasıyla ön plana çıkmıştır [9].

CNET PSOLA, 1980'li yılların ortalarında Fransa Telekom CNET (Centre National d'Etudes Télécommunications) tarafından çift-ses (diphone) ekleme yöntemi kullanılarak geliştirilmiştir. İngiliz ve Amerikan İnglizcesi, Fransızca, İspanyolca ve Almanca desteği bulunmaktadır [7].

MBROLA projesi, Belçika Faculte Polytechnique de Mons TCTS Laboratuvarlarında geliştirilmiştir ve asıl amacı çoklu dil destekli, ticari olmayan ve araştırma odaklı bir metin seslendirme uygulaması tasarlamaktır. Projede PSOLA benzeri algoritma kullanılmıştır ancak CNET patenti dolayısıyla bu isim yerine MBROLA kullanılmıştır [10].

GVZ, SESTEK firması tarafından sadece Türkçe için geliştirilmiş ticari üründür. Eklemeli yöntem kullanılarak elde edilen GVZ TTS yazılımının amacı elektronik ortamdaki metnin anlaşılır biçimde ve insan sesi doğallığında seslendirilmesidir. Türkçe için başarılı sonuçlar elde edilmiştir.

3. Türkçe Metin Seslendirme Sistemi

Türkçe için, mümkün olduğunca doğal ve anlaşılır, metinden ses üretme sisteminin gerçeklenmesi amacıyla literatürdeki çalışmalardan ikisi göz önünde bulundurulmuş, geliştirmesi yapılmıştır. Bu çalışmalara ulama, hece geçişleri v.b. gibi dilbilimsel etkenler de eklenmiştir. Ayrıca ses uzunluğu ve genlik değişimleriyle anlaşılırlığın arttırılması ve istatistiksel olarak gösterilmesi sağlanmıştır.

3.1. Metin Çözümlemesi

Öncelikle, girilen metnin söyleyişteki karşılığının elde edilmesi için çözümlemeler gerekmektedir. Bu aşama dile çok bağımlıdır ve dile özgü çözümler içermektedir.

Metin önişleme aşamasında rakamlar, sayılar, kesirler, tarihler, sıra belirten ifadeler, kısaltmalar ve özel karakterler gibi yazı dilinde anlamı olan ifadeler, okunurken sarf edilen sözcüklere dönüştürülmekdir.

Örneğin: "1876" sayısı "binsekizyüzyetmişaltı" şeklinde okunacak şekle dönüştürülmelidir. Kentilyon mertebesine kadar sayıların çevrilebilmesi sağlanmıştır.

Benzer biçimde, "4/5" kesir ifadesi "4 bölü 5" veya "beşte dört" şeklinde, "11.04.1978" veya "11/04/1978" gibi tarih ifadeleri de "onbir nisan bindokuzyüzyetmişsekiz" veya "onbir dört bindokuzyüzyetmişsekiz" şeklinde çözümlenebilmektedir.

Dilbilimsel çözümleme aşamasında sözcüğün cümle içindeki anlamına göre seslendirme yoluna gidilmelidir. Örneğin "Ayşe hala gelmedi" cümlesinde bulunan "hala" sözcüğü iki anlamda kullanılabilmektedir. Babanın kızkardeşi anlamında kullanılmışsa sert okunması gerekirken, henüz anlamında kullanılmışsa yumuşak seslendirilmelidir. Ancak bu cümlede hangi anlamda kullanıldığının tesbiti konusunda kesin bir yöntem olmadığı için bu konuda çalışma yapılmamıştır. Buna benzer bir şekilde "kağıt" sözcüğündeki 'k' sesi ile "kalmak" sözcüğündeki 'k' sesi birbirinden farklıdır. Bu gibi cümleye göre seslerin nasıl okunması gerektiğine karar verme işlemi, doğal dil işleme konularında yapılacak çalışmalarla mümkün olabilmektedir.

Ölçü çözümleme aşamasında metnin doğru vurgu ve tonlamada seslendirilebilmesi için hesaplama yapılmaktadır. Ancak çalışma dahilinde ölçü çözümleme işlemi sadece sözcük ve/veya cümlenin bulunduğu yere göre kurallar dahilinde yapılmaktadır. Örneğin her cümlenin sonundaki sözcük ile sözcüğün sonunda bulunan hece diğerlerine göre belli katsayıda daha yüksek ve uzun okunacak şekilde çözümleme yapılmaktadır.

3.2. Ses Çözümlemesi

Gerekli olan ses veri tabanının oluşturulması ve bu veri tabanında bulunan her bir ses dosyalarının etiketlenmesi oldukça uzun zamanlar almaktadır. Ayrıca, eklemeli yöntemlerde; seslerin eklenme yerlerinde gürültüler oluşabilmektedir. Bunun için sesler arası geçişlerde çeşitli algortimalar kullanılabilmektedir.

3.3. Ses Dosyalarının Hazırlanması

Ses dosyaları birçok değişik formatta saklanabilir fakat bu çalışmada en çok bilinen formatlardan biri olan "wav" kullanılmıştır. Wav dosyası üç veri bölgesi (chunk) içermektedir:

Birinci veri bölgesi olan RIFF 12 byte uzunluğundadır ve dosyanın bir "wav" dosyası olduğunun belirtildiği bölgedir. RIFF veri bölgesi alanları Çizelge-2'de gösterilmiştir.

Çizelge-2: Önemli Metinden Ses Üretme Sistemleri [11]

sekizli sırası	Açıklama
0 - 3	RIFF (ASCII karakterleri şeklinde)
4 - 7	Little Endian Şekilde paketin geri kalanının boyutu
8 - 11	WAVE (ASCII karakterleri şeklinde)

İkinci veri bölgesi FORMAT'tır. Bu bölgede formata özgü parametreler tanımlanmaktadır ve 24 byte uzunluğundadır. FORMAT veri bölgesi alanları Çizelge-3'te gösterilmiştir.

Çizelge-3: FORMAT veri bölgesi (chunk)– 24 sekizli [11]

byte sırası	Açıklama
0 - 3	RIFF "fmt" (ASCII karakterleri şeklinde)
4 - 7	FORMAT bölgesi uzunluğu (Binary, daima 0x10)
8 - 9	Daima 0x01
10 - 11	Kanal sayısı (Mono : 0x01, Streo : 0x02)
12 - 15	Hz olarak örnekleme oranı (binary)
16 - 19	Saniyedeki sekizli miktarı
20 - 21	Örnekteki sekizli anlamı : 1 = 8 bit mono, 2 = 8 bit stereo veya 16 bit mono, 4 = 16 bit stereo
22 - 23	Örnekteki bit sayısı

Üçüncü veri bölgesi ise DATA'dır ve bu alanda gerçek örnekleme verileri tutulur. DATA veri bölgesi alanları Çizelge-4'te gösterilmiştir.

Çizelge-4: DATA veri bölgesi (chunk) [11]

byte sırası	açıklama
0 - 3	"data" (ASCII karakterleri şeklinde)
4 – 7	Verinin uzunluğu
8 – son	Veri (Örnekler)

WAV dosya formatına uygun şekilde tüm ses dosyalarının kaydedilmesi ve üretilmesi; oldukça dikkatli ve titiz yapılması gereken bir aşamadır. Öncelikli olarak TDK sesli sözlük veri tabanı incelenerek, çift-ses ekleme (diphone concatenation) ve hece ekleme (syllable concatenation) yöntemleri test edilmiştir. Eklemeli yöntemlerde sesler mümkün olduğunca tekdüze ve ritimsiz olmalıdır. Ancak; TDK sesli sözlük veri tabanındaki sesler arasında erkek ve kadın sesleri karışık olarak yer alması, seslerin farklı vurgu ve tonlamayla seslendirilmiş olması dolayısıyla bu sesli sözlük kullanılmamıştır.

Her bir çift-ses ve hecenin önceden kaydedilmesi ve etiketlenmesi çok uzun bir süreçtir. Bu yüzden; öncelikle, küçük bir veri tabanı oluşturulması yoluna gidilmiştir. Birkaç cümle için başarılı sonuçlar elde edilince, çalışmanın kapsamı büyütülüp tüm çift-ses ve heceleri içermesi hedeflenmiştir. Ancak kısıtlı zaman içinde tüm zamanın ses kaydedilmesi ve etiketlenmesi ile uğraşmak yerine, yazılacak bir program vasıtasıyla otomatik olarak üretilmesi düşünülmüştür.

Çalışmanın hedeflerinden birisi de, ses kaydının titiz bir çalışma sonucunda, düzgün ve monoton kaydedilmesi ile, sonucun da anlaşılır ve doğal olabileceğinin gösterilmesi olduğundan MBROLA ile tüm ses dosyalarının otomatik olarak üretilmesi ve sınanması sağlanmıştır. Mbrola ile ses üretimi yapılabilmesi için fonetik işaret gerekmektedir. Mbrola'nın desteklediği ve desteklemediği sesleri içeren Türkçe sesçil alfabesi Çizelge-5'te gösterilmiştir. Bu alfabe oluşturulurken çizelgede SAMPA ve MBROLA'nın kullanmış olduğu fonetik işaretler de dikkate alınmıştır [12]. Bazı sesler MBROLA veri tabanında yer almamaktadır (Çizelgede MBROLA sütununda "-" olarak belirtilmişlerdir). Çizelgeye ayrıca eklenecek dosya adının saptanabilmesi için sese ait dosya adı karakteri de eklenmiştir.

Çizelge-5: Türkçe Sesçil Alfabesi

Harf	Örnek sözcük	SAMPA	MBROLA	Dosya adı
a	kal, aşk	a	a	A
b	balık, batak	b	b	b
c	cam, can	dZ	dZ	c
ç	seçim, çan	tS	tS	c2
d	dede, dudak	d	d	d
e	keçi, yemek	e	e	e
f	fakat, fare	f	f	f
g	geri, gemi,	gj	g	g
g	karga, gaga	g	-	-
ğ	sağ (sol tersi),	G	G	g2
h	hasta, hasan	h	h	h
ı	kıl, sınav	1	@	i2
i	kil, izin	i	I	i
j	müjde, jeton	Z	Z	j
k	akıl, kalın	k	k	k
k	kedi, keser	c	-	-
l	pala, sal	5	l	l
l	lale, lavanta	l	L	l2
m	dam, maymun	m	m	m
n	anı, nasıl	n	n	n
n	süngü, düğün	N	-	-
o	kol, osman	o	o	o
ö	göl, ölü	2	@	o2
p	ip, para	p	p	P
r	raf, para	r	r	r
s	ses, sakat	s	s	s
ş	aşı, kaş	S	S	s2
t	ütü, tarak	t	tS	t
u	kul, usta	u	u	u
ü	kül, ürkek	y	y	u2
v	ver, kavak	v	v	v
v	tavuk	w	-	-
y	yat, kayak	j	j	y
z	azık, kazak	z	z	z

Türkçede Bulunan Hece Tipleri

Türkçe sesçil abecesi dikkate alındığında üretilmesi gereken sekiz farklı hece türü bulunmaktadır. Bunlar en az bir, en fazla dört harften oluşur. Aslında öz Türkçe'de altı farklı hece tipi bulunmaktadır, ancak günümüzde diğer dillerden gelen ve dilimizce benimsenmiş birçok sözcük bulunmaktadır ve bu hece tipleri dahil edilmediği takdirde çoğu metin seslendirmesinde sorun yaşanacaktır. Bu yüzden çalışma kapsamına yabancı kökenli sözcüklerde görülen hecelerin büyük çoğunluğu da eklenmiştir.

Hece tiplerini belirlerken "C" sessiz, "V" sesli harfleri belirtmek üzere kullanılacaktır.

- **V** tipinde heceler

Tek sesli harften oluşan hecelerdir ve toplamda sekiz adet V tipinde hece bulunmaktadır : (*a, e, ı, i, o, ö, u, ü*)

- **CV** tipinde heceler

Sessiz harf + sesli harf şeklinde oluşan hecelerdir (Örnek: al, an, et, üç, öl).
Matematiksel olarak 21 x 8 = 168 adet CV tipinde hece bulunmaktadır. Ancak Çizelge 5'te Mbrola'nın desteklediği 22 adet sessiz olduğu için 22 x 8 = 176 adet ses dosyası oluşturulabilmektedir.

- **VC** tipinde heceler

Sesli harf + sessiz harf şeklinde oluşan hecelerdir (Örnek: ba, ce, zi, ğü). CV hece tipinde olduğu gibi 8 x 22 = 176 adet ses dosyası oluşturulabilmektedir.

- **VCC** tipinde heceler

Sesli harf + sessiz harf + sessiz harf şeklinde oluşan hecelerdir (ilk, ürk, ast).
Matematiksel olarak 8 x 22 x 22 = 3872 adet ses dosyası oluşması gerekmektedir. Ancak Türkçede şöyle bir kural vardır : "*Aynı hecede iki ünsüz harf varsa bu ünsüz harf çifti "lç, lk, lp,lt, nç, nk, nt, rç, rk, rp, rs, rt, st, şt" olmalıdır*" [13]. Ayrıca, yabancı kökenli sözcüklerde de bulunabilen "rz" sessizleri (örnek: tarz, ırz) ve daha çok ünlem içeren sesleniş kalıplarında görülen "yt" sessizleri (örnek: "heyt") de eklendiği takdirde Çizelge-6'daki görüntü ortaya çıkmaktadır.

Çizelge-6: Hece sonunda bulunabilen iki sessiz

Hece sonunda çift ünsüz	Örnekler
lç, lk, lp, lt	felç, kalk, alp, alt
nç,nd,nk,nt	genç,trend,denk,kent
rç,rf,rk,rp,rs,rt, rz	sürç,örf,kürk,turp,hırs,sırt,ırz
St	üst
Şt	Rüşt
Yt	Hayt

Ayrıca "–l" sessizi, "lale" ve "halı" sözcüklerinde farklı seslendirildiği için bu kuralda "-l" ile ilgili sesler de çoklanmalıdır.
8 x 14 = 112 (kurala göre oluşması gereken ses dosyaları toplamı)

8 x 1 = 8 ("yt" ile biten ses dosyaları toplamı)
8 x 4 = 32 (-lç, -lk, -lp, -lt ile biten ve "lale" sözcüğündeki "l" sesinin kullanıldığı ses dosyaları toplamı)

olmak üzere toplam 152 ses dosyası bulunmaktadır.

- **CVC** tipinde heceler

Sessiz harf + sesli harf + sessiz harf şeklinde oluşan hecelerdir (örnek: kal, tek, bit).
22 x 8 x 22 = 3872 adet ses dosyası gerekmektedir.

- **CVCC** tipinde heceler

Sessiz harf + sesli harf + sessiz harf + sessiz harf şeklinde oluşan hecelerdir (türk, sark, dört). VCC tipindeki hecelere ait ses dosyası sayısı hesabına benzer olarak :

22 x 8 x 14 = 2464 (kurala göre oluşması gereken ses dosyası toplamı)

22 x 8 x 1 = 176 ("yt" ile biten ses dosyaları toplamı)

22 x 8 x 4 = 704 (-lç, -lk, -lp, -lt ile biten ve "lale" sözcüğündeki "l" sesinin kullanıldığı ses dosyaları toplamı)

olmak üzere toplam 3342 adet ses dosyası bulunmaktadır.

- **CCV** tipinde heceler

Sessiz harf + sessiz harf + sesli harf şeklinde oluşan hecelerdir (örnek: tra, spo, gri). Yabancı kökenli sözcüklerde bulunabilen CCV tipi hece yapısı, sadece sözcük başlarında bulunabilir. Matematiksel olarak 22 x 22 x 8 = 3872 adet ses dosyası oluşturulmalıdır ancak yabancı kökenli sözcüklerin başında bulunabilen bu hece tipi sadece Çizelge 7'de görüldüğü üzere (br, bl, dr, fr, gl, gr, hr, kl, kr, pl, ps, tr) listesindeki iki sessizlerle başlayabilirler. Dolayısıyla CCV hece tipinde 12 x 8 = 96 adet ses dosyası gerekmektedir.

Çizelge-7: Hece başında bulunabilen iki sessizler

Hece başında çift ünsüz	Örnekler
bl, br, dr, fr	blok, briç, draje,drenaj, fransız, fren
gl, gr	glikoz, gram
hr,kl,kr	hristiyan, klor, krom
pr, ps	pranga, psiko
Tr	tren, troleybüs

- **CCVC** tipinde heceler

CCV hece tipine benzer bir şekilde yabancı kökenli sözcüklerde görülmektedir ve sadece sözcüğün başında bulunabilmektedir. Yine CCV tipindeki heceler gibi (br, bl, dr, fr, gl, gr, hr, kl, kr, pl, ps, tr) iki sessizleriyle başlayabilirler.

Hem CCV hem de CCVC tipindeki heceler konuşma dilinde Türkçe'nin temel 6 hece yapısına indirgenebilir. İndirgeme sonucu CCV türü heceler CV-CV olacak şekilde (örneğin "trafo" sözcüğünde bulunan "tra" hecesi "tı-ra" şeklinde) , CCVC türü heceler CV-CVC olacak şekilde (örneğin "tren" hecesi "ti-ren" şeklinde) iki farklı heceye dönüştürülebilir. CCV tipindeki hecelerin sayısı sadece 96 iken CCVC tipindeki heceler 12 x 8 x 22 = 2112 adet ses dosyası gerektirmektedir. Dolayısıyla CCVC tipindeki hecelerde bu indirgeme aktif hale getirilmiş, CCV tipindeki hecelerde ses üretme yoluna gidilmiştir.

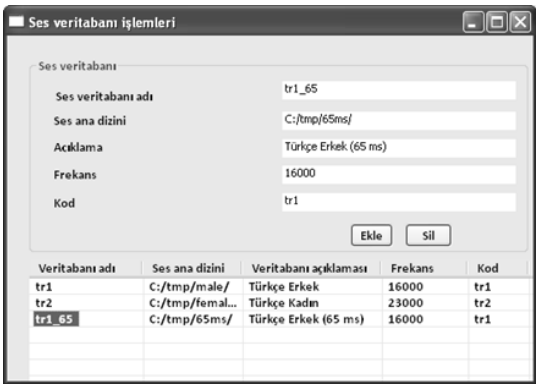
Belirtilen tüm hece tipleri gözönünde bulundurulduğunda, hece eklemeli yöntemde, Türkçe metin seslendirme sistemi için yaratılması gereken ses dosyaları toplamı (en çok) Çizelge-8'de gösterilmiştir.

Çizelge-8: Türkçe için oluşturulması gereken ses dosyaları toplamı

	Hece yapısı	Örnek	Matematiksel ses dosyası toplamı	Kurallar ile ses dosyası toplamı
1	V	a, e, ı, i, o, ö, u, ü	8	8
2	CV	al, an, et, üç, öl	21 x 8 = 168	22 x 8 = 176
3	VC	ba, ce, zi, ğü	8 x 21 = 168	8 x 22 = 176
4	VCC	ilk, ürk, ast	8 x 22 x 22 = 3872	8 x 14 8 x 1 8 x 4 = 152
5	CVC	kal, tek, bit	22 x 8 x 22 = 3872	22 x 8 x 22 = 3872
6	CVCC	türk, sark, dört	22 x 8 x 22 x 22 = 85184	22 x 8 x 14 22 x 8 x 1 22 x 8 x 4 = 3342
7	CCV	tra, spo, gri	22 x 22 x 8 = 3872	12 x 8 = 96
8	CCVC	tren, kral	22 x 22 x 8 x 22 = 85184	Hece indirgeme ile 0
		Toplam	**182328**	**7822**

Heceler oluşturulurken dikkat edilmesi gereken noktalardan birisi hece uzunluğudur. Heceler veya çift-sesler (diphone) sözcüğün başında, ortasında veya sonunda bulunma durumlarına göre farklı uzunlukta olabilirler. Bu yüzden, oluşturulan heceler, sözcüklerin oluşturulmasında da kullanılacağı için ortalama uzunlukta heceler seçilmiştir. Bu yüzden, normal konuşmada her ses için ortalama süre 65 ms olarak ele alınmış, tek kanal (mono), 44100 örnekleme oranı, PCM (darbe kod modülasyonu) ve 16 bitlik örnekler olacak şekilde oluşturulmuştur.

Girdi olarak alınan metnin değişik ses veritabanları tarafından seslendirilebilmesine imkan verilmektedir. Üretilen ses dosyalarını içeren ses veri tabanı tanımlaması Şekil-3'te görülen ekran aracılığıyla tanımlanabilmektedir.



Şekil-3: Ses veri tabanının uygulamaya tanıtılması

3.4. Oluşturulan Seslerin Genliklerinin Dengelenmesi

Otomatik olarak oluşturulan ses dosyaları arasında dengelenmemiş (farklı genlikte) ses dosyaları mevcut olabileceğinden ekleme yerlerinde seste çatlamalar oluşmaması amacıyla, tüm dosyalar üzerinde basit bir algoritma çalıştırılarak seslerin dengelenmesi hedeflenmiştir. Algoritma basitçe şu şekilde çalışmaktadır :

Öncelikle, tüm ses dosyaları taranarak, en yüksek, en düşük ve ortalama genlik değerleri belirlenir. Tüm ses dosyaları ikinci kez tarandığında, bir üst adımda belirlenmiş olan en yüksek, en düşük ve ortalama genlik değerlerine göre karşılaştırılarak genlik değerleri aşağıdaki üç yöntemle şekilde dengelenebilir.

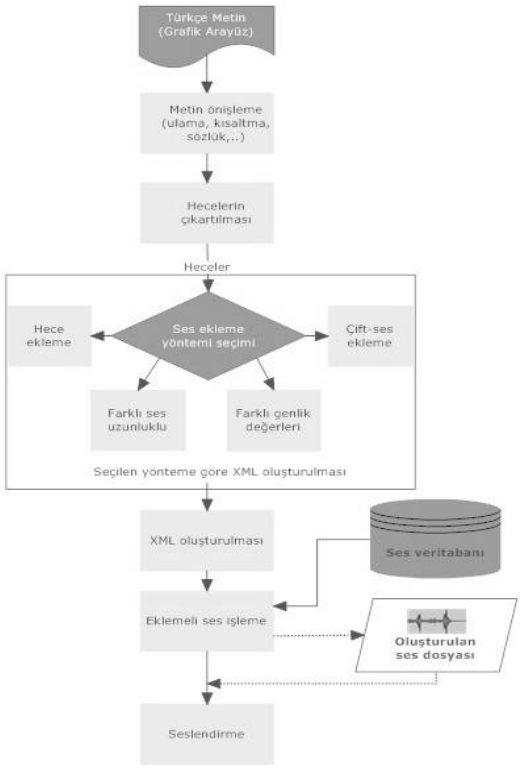
1. En düşük genlik değerine göre dengeleme: Tüm seslerdeki en düşük genlik değeri ile o an işlenecek olan ses dosyasının en düşük genlik değeri karşılaştırılıp aralarındaki oran doğrultusunda tüm genlik değerlerinin yeniden ayarlanması.
2. En yüksek genlik değerine göre dengeleme: Tüm seslerdeki en yüksek genlik değeri ile o an işlenecek olan ses dosyasının en yüksek genlik değeri karşılaştırılıp aralarındaki oran doğrultusunda tüm genlik değerlerinin yeniden ayarlanması.
3. Hem en yüksek, hem de en düşük genlik değerine göre (ortalama) dengeleme: Tüm seslerdeki ortalama genlik değeri ile o an işlenecek olan ses dosyasının ortalama

genlik değeri karşılaştırılıp aralarındaki oran doğrultusunda tüm genlik değerlerinin yeniden ayarlanması.

Dengeleme çalışması kısmen başarılı olmuştur ancak istatistiksel olarak başarı oranı tespit edilmemiştir.

3.5. Türkçe Metin Seslendirme Sisteminin Gerçeklenmesi

Şekil-4'te gerçeklenen Türkçe Metin Seslendirme sisteminin temel akış şeması gösterilmektedir.



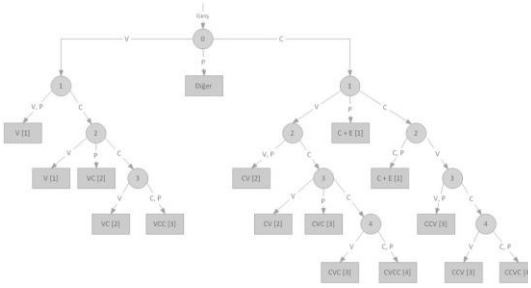
Şekil-4: Türkçe Metin Seslendirme Sisteminin temel akış şeması

3.5.1 Metin İşleme Bileşeni

Sisteme girdi olarak gelen metin, içinde bulunan noktalama işaretlerine ve boşluklara göre ağaç yapısına dönüştürülür. Bu ağaç yapısında metin, cümle, sözcük, hece, ses ve noktalama işaretleri bulunmaktadır. Her bir cümle içinde bulunan sözcükler hecelerine parçalanır. Bunun için hece parçalama algoritması çalıştırılır. Bu algoritma öncesinde "ulama" seçimi yapıldıysa, cümlenin ulamalı şekilde hecelere ayrılması sağlanır. Ulama seçeneği, konuşma işleminin anlaşılır ve doğal olması için kilometre taşlarından birisidir. Örneğin, "Yeşil ağacın altında uzanıyordu." cümlesini "Ye-şil a-ğa-cın al-tın-da u-za-nı-yor-du" şeklinde hecelere ayırabiliriz, ancak ulama seçeneği ile bu hecelere ayırma işlemi şu şekilde olmaktadır :
"ye-şi-la-ğa-cı-nal-tın-da u-za-nı-yor-du".
Cümle bu şekilde okunduğunda, daha doğal olmaktadır.

Sözcüğün hecelerini çıkarma algoritması Şekil-5'te görüldüğü şekilde çalışır. Bu ağaçta her düğümden üç tane kol çıkmakta olup ünlü (V), ünsüz (C) ve bunların dışındaki karakterler (P) ile gösterilmektedir. İnceleme yukarıdan aşağıya doğru yapılmaktadır. Her düğümde bulunan halkaların içindeki rakamlar, incelenen karakterin, hece başından itibaren kaçıncı karakter olduğunu göstermektedir. Metnin incelenen kısmı içinde ilk karakterden başlayarak her karakterin simgelediği dal sırasıyla takip edilirse, sonuçta o karakterle başlayan heceye ait son düğüme ulaşılacaktır. Bu düğümde, içinde hece türü ve bir sonraki hecenin başlangıcına ulaşmak için tarama işlemine kaç karakter öteden devam edileceği bilgisi bulunan kutular vardır. Tarama işlemi, metnin sonunu simgeleyen karaktere ulaşılıncaya kadar devam eder.

Ağaç yapısının daha iyi anlaşılması için "ödev" sözcüğünün heceleri, ağacı izleyerek bulunabilir. Sözcüğün, dolayısıyla ilk hecenin, birinci karakteri bir ünlü olan "ö" dür. Yani, hece başlangıcı olan 0. karakter bir ünlüdür ve soldaki "V" koluna dallanmak gereklidir. Hecenin birinci karakteri "d" bir ünsüz olup bu sefer, birinci düğümün altında sağdaki "C" koluna dallanmalıdır. Bundan sonra, hecenin ikinci karakteri olan "e" ünlüsü için ikinci düğümün altından sol kola geçilir. Burada ağaç sonlanmış, sözcüğün ilk hecesinin "V" türünde olduğu anlaşılmıştır. "[1]" bilgisi ile, sözcüğün ele alınacak yeni hecesinin ilk karakterinin "ö" den sonraki birinci karakter ("d") olduğu anlaşılmıştır.



Şekil-5: Hecelere parçalama algoritması

(C + E [1]) yapısına, yanında bir ünlü olmayan ünsüze rastlandığında veya art arda gelen ünsüzlerden anlam çıkartılamadığında gelinir. Örneğin, "spor" sözcüğündeki "sp" ünsüzleri, bunları takip eden "o" ünlüsü nedeniyle, CCV ya da CCVC yapısına uygun oldukları için, anlam taşımalarına rağmen "PTT" şeklindeki ünsüz dizileri hecesel anlam taşımazlar. Heceleme ağacına bu durumun da eklenmesinin asıl nedeni, "n tane tamsayı" şeklinde, ünsüzün tek başına kullanıldığı ve "PTT kurumu" şeklinde, içinde kısaltmalar olan cümlelerle sık sık karşılaşılmasıdır. Böyle bir durumla karşılaşıldığında, ünsüzün yanına "E" ünlüsü eklenerek " C + 'E' " şeklindeki hecenin seslendirilmesi yoluna gidilir. Bilindiği gibi Türk alfabesindeki tüm ünsüzler, yanlarında 'E' ünlüsü varmış gibi seslendirilirler (B → BE, D → DE gibi). Bu durumda yukarıdaki sözcükler hecelere ayrılma aşamasından sonra "ne ta-ne tam-sa-yı" ve "Pe Te Te ku-ru-mu" şeklini alır.

Hece parçalama algoritması sonrasında ana programda Şekil-6'daki gibi heceler farklı biçimlerde listelenebilir.

Şekil-6: Hece parçalama algortiması sonrasında görünüm

Metin işleme bileşeninin son aşaması Şekil-7'de görülen XML katarının üretilmesidir.

```
<?xml version="1.0" encoding="ISO-8859-
9"?>
<metin>
   <cumle>
     <kelime vurguKatsayisi="1">
       <hece vurguKatsayisi="1">
          <ses sampa="j" sure="60"/>
          <ses sampa="e" sure="90"/>
       </hece>
       <hece vurguKatsayisi="1">
          <ses sampa="S" sure="60"/>
          <ses sampa="I" sure="90"/>
       </hece>
       <hece vurguKatsayisi="1">
          <ses sampa="l" sure="60"/>
          <ses sampa="a" sure="90"/>
       </hece>
       …
       <hece vurguKatsayisi="1.4">
          <ses sampa="d" sure="90"/>
          <ses sampa="u" sure="120"/>
       </hece>
     </kelime>
     <bekle sure="50"/>
   </cumle>
   <bekle sure="100">.</bekle>
</metin>
```

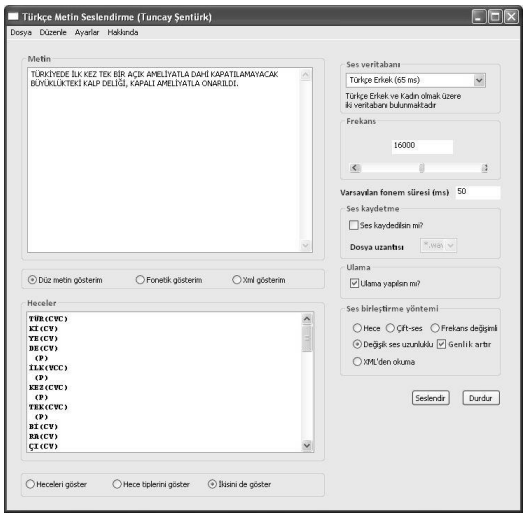
Şekil-7: Metin işleme biriminde oluşturulan XML örneği

3.5.2 Ses İşleme Bileşeni

Metin işleme biriminin oluşturmuş olduğu XML katarı, ses işleme bileşeninde sese dönüştürülür. Dönüştürme işlemi dört farklı yöntem ile yapılabilmektedir.

1. Çift-ses eklemeli yöntem
2. Hece eklemeli yöntem
3. Farklı hece uzunlukları ile eklemeli yöntem
4. Farklı genlik değerleri ile eklemeli yöntem

Ayrıca her bir yöntem için "ulama" seçeneği de isteğe bağlı olarak eklenebilmektedir. Her bir yöntem ve ulama seçeneği, Şekil-8'de görüldüğü üzere, seçilerek çalıştırılabilir şekilde tasarlanmıştır.



Şekil-8: Türkçe Metin Seslendirme uygulaması ekran görüntüsü

Anlaşılır konuşma frekans bandının 5 kHz ve örnekleme frekansının, örnekleme teoremi uyarınca, 16 kHz olduğu ortalama kayıt süresi 200 ms olan 16 bitlik yaklaşık 8000 hecenin, ortalama 1 saniyelik kaydı için yaklaşık 50MB saklama alanı gerekmektedir (2x16000x0.2x8000/(1024x1024)). Günümüzde bu değer önemsiz sayılsa da, geçmiş yıllarda önemli kısıtlardan biri olarak hesap edilmekteydi. Bu yüzden çift-ses ekleme yöntemi daha çok revaçtaydı. Ayrıca, hece sayısının fazla olması nedeniyle hazırlık süresinin ve heceler arasındaki normalizasyon sorununun da çıkması çift-ses yöntemini daha tercih edilir hale getirmiştir. Sonuç olarak, hece eklemeli yöntem ile kıyaslanması, hem de hece ses veri tabanı hazırlama işleminin oldukça uzun uğraşlar gerektirmesi dolayısıyla çift-ses eklemeli yöntem de çalışmaya dahil edilmiştir.

Girilen metnin çift-sesler ile seslendirilebilmesi için öncelikle metin içinde bulunan hecelerin çift-sesler birlikteliğine dönüştürülmesi gerekmektedir. Bunun için Çizelge-9 hazırlanmıştır.

Uygulamada, girilen metne ait seslendirme yapılabilmesi için öncelikle Şekil-8'de görülen "Ses birleştirme yöntemi" panelinden bir yöntem seçilmelidir.

Çizelge-9: Hecelerin uygun çift-seslere bölünmesi

Hece tipi	Uygun çift-sesler
(V)	V + (bir sonraki hece C ile başlıyor ise bekle)
(VC)	V + VC
(CV)	CV + V + (bir sonraki hece C ile başlıyor ise bekle)
(VCC)	V + VC + C
(CVC)	CV + V + VC + C
(CVCC)	CV + V + VC + C

Çift-ses eklemeli yöntemde olduğu gibi, hece eklemeli yöntemde de her ses sabit uzunlukta ele alınmıştır. Yapılan deneyler sonrasında en anlaşılır şekilde konuşmanın üretildiği ses uzunluğu 65 ms olarak belirlenmiştir. Bu yüzden üretilmiş ve kaydedilmiş veritabanlarından "Türkçe Erkek (65ms)" kullanılmıştır. Bunun yanında "Türkçe Kadın (65 ms)", "Türkçe Erkek (55 ms)" ve "Türkçe Kadın (55 ms)" ses veri tabanları da bulunmaktadır.

Her iki yöntemde de tüm ses uzunluklarının sabit olması dolayısıyla, anlaşılır fakat doğallıktan uzak sonuçlar ortaya çıkmıştır. Doğallığı arttırabilmek için üzerinde durulması gereken konulardan ikisi vurgu ve uzatmalardır. Farklı hece uzunlukları ile eklemeli yöntemde ses sürelerinde uzatmalar yapılarak doğallığı arttırma yoluna gidilmiştir. Belirlenen genel kurala göre hecelerde bulunan sessizler 60 ms, sesliler ise 90 ms olacak şekilde tanımlanmıştır. Ancak hece, sözcüğün son hecesi ise sessizler için 90 ms, sesliler için ise 120 ms esas olarak alınmıştır. Ses analizi bu değerlere göre yapılacak şekilde XML katarı üretilmiştir. Örnek olarak Şekil-7'de üretilen XML katarı (her ses öğesi için "sure" parametresi üretilmektedir) incelenebilir.

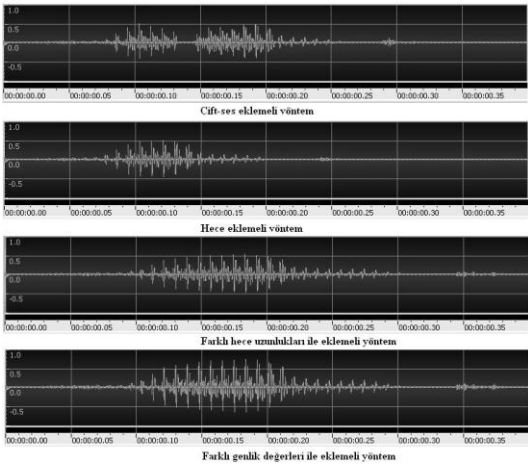
Türkçe sözcüklerde, sözcük içinde ve cümle içinde vurgunun nerede yapılacağı yaklaşık olarak bellidir; ancak konuşma dilinde vurgunun yeri değişebilir [14]. Bu çalışmada sadece belli formül altında denenmeye çalışılmış ve asıl çalışmalar doğal dil işleme bileşenine bırakılmıştır. Türkçe Metin Seslendirme sisteminde vurgu, genlik değişimi ile verilmeye çalışılmıştır. Ses uzunluğunun arttırıldığı formülle benzer şekilde, ses genliği de belli katsayı ile arttırılarak sesin daha şiddetli çıkması sağlanmıştır. Bu da cümlenin geneline bakıldığında

doğallığı arttırıcı sonuçlar vermiştir. Yine Şekil-7'de görülen XML örneğinde "vurguKatsayisi" öğesi ile genlik katsayısı belirlenmektedir.

Tüm yöntemlerin yanı sıra, bir de doğrudan XML'den okuma seçeneği eklenmiştir. Bu seçenek, eklenecek olan doğal dil işleme bileşeninin üreteceği XML'i doğrudan sese dönüştürebilecek şekilde tasarlanmıştır. Şekil-7'de örnek XML'de gösterildiği gibi aşağıdaki özellikler bulunabilmektedir.

- Metin (<metin>) içinde bulunacak duraksamalar (ms cinsinden), noktalama işaretleri (<bekle>) ile belirtilir.
- Cümle (<cumle>) içinde bulunan tüm sözcükler (<kelime>), ve bu sözcükler arası duraksamalar (<bekle>), sözcüklere ait vurgular ("vurguKatsayisi") ile belirtilir.
- Sözcükler içinde bulunan hecelere ait vurgular ("vurguKatsayisi") ile belirtilir.
- Hecelerin içinde bulunan tüm seslere (<ses>) ait ses özel işareti ("sampa") ve ms cinsinden süre bilgisi ("sure") ile belirtilir.

Şekil-8'de "halk" sözcüğüne ait, her bir yöntem ile oluşturulmuş seslerin dalga şekilleri gösterilmiştir.



Şekil 8: "halk" sözcüğünün üretilmiş ses dosyalarına ait dalga şekilleri

4. Sistemin Değerlendirilmesi

Ses kalitesini değerlendirmek için dünyada kullanılan en yaygın ve basit yöntem MOS (Mean Opinion Score)'tur. MOS'ta, 1-kötü ve 5-mükemmel arasında beş farklı seviye vardır ve bu seviyeler Çizelge-10'da listelenmiştir. Dinleyiciler, dinledikleri sesleri değerlendirerek beş seviyeden birini uygun görürler [1].

Çizelge-10: MOS Seviyeleri

MOS seviyeleri		
5	Mükemmel	Excellent
4	İyi	Good
3	Normal	Fair
2	Zayıf	Poor
1	Kötü	Bad

Ancak, dinleyicilerin sadece beş seviyeden birini seçmesi çoğu zaman anlamlı sonuçlar veremeyebilir. Özellikle seviyeler dinleyiciden dinleyiciye değişkenlik gösterebilir. Bu yüzden değerlendirmenin neye göre ve nasıl yapılacağı üzerine bir çalışma yapılmıştır. Dinleyici seslendirilen cümleyi doğru olarak tekrar edememişse, cümle tekrar seslendirilir ve bu işlem en fazla üç kere tekrar edilir.

Değerlendirmeye ve dolayısıyla anlaşılırlığın hesaplanmasına dahil olması gereken parametreler aşağıda özetlenmiştir.

doğru harf sayısı: dinleyicinin anladığı cümle ve gerçek cümledeki çakışan harf sayısını gösterir. (Anlaşılırlık ile doğru orantılı)

önceki doğru harf sayısı: Bir önceki tahminde bulunan doğru harf sayısı. İlk tahmin için bu değer 0 (sıfır)'dır. (Doğru harf sayısı ile farkı, anlaşılırlık ile doğru orantılı)

toplam harf sayısı: Seslendirilen cümlede bulunan harf sayısı (Doğru bilinen harf sayısı ile bereber düşünüldüğünde, sonuç ile ters orantılı)

deneme numarası (n) : Aynı cümle için kaçıncı denemenin olduğunu gösterir. (Cümle llk seferde doğru bir şekilde bilinemediyse, diğer denemelerde verilen cevapların değeri daha az olmalıdır, bu yüzden deneme numarası, sonuç ile ters orantılıdır)

Tüm parametreler göz önüne alındığında anlaşılırlık formülü Şekil 9'daki gibi elde edilmiştir.

$$p(x) = \sum_{n=1}^{3} \left(\frac{doğru\ harf\ sayısı - önceki\ doğru\ harf\ sayısı}{toplam\ harf\ sayısı} \frac{1}{n} \right)$$

Şekil-9: Anlaşılırlık formülü

Örnek olarak, "*Çok fazla kar yağdığı için annem işe gidemedi*" cümlesi dinleyiciye dinletilmiş ve üç denemede şu sonuçlar alınmış olsun:

1. Deneme: Çok fazla kar vardı anne işe gidemedi. (*29 harf doğru*)

2. Deneme: Çok fazla kar yağdı annem işe gidemedi. (*31 harf doğru*)

3. Deneme: Çok fazla kar yağdığı için annem işe gidemedi. (*38 harf doğru*)

Asıl cümlede 38 harf bulunmaktadır ve yukarıdaki değerler formüle yerleştirildiğinde birinci denemeden gelen sonuç %76,32 iken ikinci ve üçüncü denemelerden gelen sonuçlar %2,63 ve %6,14 olmaktadır. Toplamda ise %85,09 değeri elde edilmiştir.

4.1 Cümlelerin Belirlenmesi

Test sonucunun daha anlamlı olması için, yoruma imkan vermeyen ve her yaştaki insana hitap edebilecek (üç ve yedi yaşında dinleyicilerin de olduğu düşünülerek) net cümleler seçilmeye gayret edilmiştir. Seçilen cümleler Çizelge-11'de gösterilmektedir.

Çizelge-11: Dinleyicilere dinletilmek üzere hazırlanan 10 cümle

Sıra Numarası	Cümleler
1	Bu sabah erken kalktım
2	Bu akşam çok yemek yedim
3	Çok fazla kar yağdığı için annem işe gidemedi
4	Parkta oynayan çocuklar uçurtma uçurdu
5	Akşam yatmadan önce süt içerim
6	Televizyonda çizgi film seyretmeyi çok severim
7	Babamla futbol oynadık
8	Artık yatma vakti geldi
9	Polis hırsızı yakaladı
10	Dişlerinizi her gün iki kere fırçalamalısınız

4.2 Cümlelerin Dinletilmesi

Belirlenen her cümle dinleyicilere aynı ortamda ve farklı zamanlarda dinletilmiştir. Tüm yöntemlerin test edilebilmesi amacıyla, daha önceden dinletilmiş yönteme ait cümlenin hatırlanmaması için bir ay gibi bir sürenin beklenmesi öngörülmüştür. Bu sürenin de yetersiz olabileceğinden yola çıkarak, her dinleyicide denenen yöntemlerin sırası karışık olacak şekilde ayarlanmıştır. Örneğin, birinci dinleyiciye ilk olarak çift-ses eklemeli yöntem ile üretilen sesler dinletilmiş iken, ikinci dinleyiciye ilk olarak hece eklemeli yöntem dinletilmiştir. Aynı şekilde üçüncü dinleyiciye de farklı uzunluklu ses ve farklı genlik değerleri içeren yöntem ile üretilen sesler dinletilmiştir. Bir ay gibi bir süre sonra ise, her dinleyici için diğer yöntemlere geçilmiştir. Sonuç olarak her yöntemin aynı sayıda birinci, ikinci ve üçüncü olarak dinleyicilere dinletilmesine gayret edilmiştir.

4.2 Sonuçların Değerlendirilmesi

Altı farklı yaş grubundaki dinleyicilerle yapılan deney neticeleri sonucunda, her bir ses işleme yöntemin yüzde cinsinden ortalama not dağılımı Çizelge-12'deki gibi hesaplanmıştır.

Çift-ses eklemeli yöntem kullanılarak yapılan deneyde anlaşılırlık oranı %91.5 iken bu değer hece eklemeli yöntemde %96.1'e yükselmiştir. Genlik ve ses uzunluğu değişimi ile vurgu çalışmasında ise anlaşılırlık %98 olarak ölçülmüştür.

Çizelge-12: Ses işleme yöntemlerinin not dağılımına göre genel ortalamaları

Çift-ses eklemeli yöntem (65ms)	**91,5**
Hece eklemeli yöntem (65ms)	**96,16**
Vurgulu (ses uzunluk, genlik)	**98,13**

Genel ortalama dışında, her bir dinleyicinin tüm cümleler için almış olduğu not ortalamalarına ilişkin çalışma Çizelge-13'te gösterilmiştir.

Sırasıyla çift ses eklemeli, hece eklemeli ve vurgulu (ses uzunluğu ve genlik değişimi ile) yöntemler değerlendirildiğinde, anlaşılırlığın arttığı hemen tüm cümleler için, yöntemlerde sağa doğru geçildikçe anlaşılırlık notunun arttığı açıkça gözlemlenmiştir.

Çizelge-13: Dinleyici cevaplarına göre oluşan not dağılımı

No	Cümle	Dinleyiciler	Sesbirim (diphone) birleştirme yöntemi (sabit 65 ms)	Hece birleştirme yöntemi (sabit 65 ms)	Vurgulu (ses uzunluk, genlik)
1	Bu sabah erken kalktım	Batu (3 yaş)	86,84	100	100
		Arda (7 yaş)	100	100	100
		Bahar (15 yaş)	100	100	100
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	100	100	100
		Mebrure (60 yaş)	100	100	100
		Ortalama	97,80	100	100
2	Bu akşam çok yemek yedim	Batu (3 yaş)	100	100	100
		Arda (7 yaş)	100	100	100
		Bahar (15 yaş)	92,5	100	100
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	100	100	100
		Mebrure (60 yaş)	100	100	100
		Ortalama	98,75	100	100
3	Çok fazla kar yağdığı için annem işe gidemedi	Batu (3 yaş)	77,63	76,31	77,63
		Arda (7 yaş)	16,22	22,36	77,63
		Bahar (15 yaş)	82,89	92,10	82,89
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	58,77	91,22	98,68
		Mebrure (60 yaş)	82,01	98,68	100
		Ortalama	69,59	80,11	89,47
4	Parkta oynayan çocuklar uçurtma uçurdu	Batu (3 yaş)	66,66	55,88	100
		Arda (7 yaş)	23,52	100	67,15
		Bahar (15 yaş)	77,94	89,70	100
		Tarık (32 yaş)	95,58	100	100
		Pınar (34 yaş)	98,52	100	100
		Mebrure (60 yaş)	91,17	97,05	100
		Ortalama	75,57	90,44	94,52
5	Akşam yatmadan önce süt içerim	Batu (3 yaş)	83,97	100	100
		Arda (7 yaş)	100	100	100
		Bahar (15 yaş)	100	100	100
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	94,23	100	100
		Mebrure (60 yaş)	98,07	100	100
		Ortalama	96,04	100	100
6	Televizyonda çizgi film seyretmeyi çok severim	Batu (3 yaş)	68,69	93,90	100
		Arda (7 yaş)	100	100	100
		Bahar (15 yaş)	100	100	100
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	100	100	100
		Mebrure (60 yaş)	100	100	100
		Ortalama	94,78	98,98	100
7	Babamla futbol oynadık	Batu (3 yaş)	100	87,5	100
		Arda (7 yaş)	100	100	100
		Bahar (15 yaş)	100	100	100
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	100	100	100
		Mebrure (60 yaş)	100	100	100
		Ortalama	100	97,91	100
8	Artık yatma vakti geldi	Batu (3 yaş)	100	100	100
		Arda (7 yaş)	87,5	100	100
		Bahar (15 yaş)	90	100	100
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	100	100	100
		Mebrure (60 yaş)	90	100	100
		Ortalama	94,58	100	100
9	Polis hırsızı yakaladı	Batu (3 yaş)	100	100	100
		Arda (7 yaş)	100	100	100
		Bahar (15 yaş)	100	100	100
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	100	100	100
		Mebrure (60 yaş)	100	100	100
		Ortalama	100	100	100
10	Dişlerinizi her gün iki kere fırçalamalısınız	Batu (3 yaş)	83,75	84,16	92,5
		Arda (7 yaş)	56,25	80,83	91,25
		Bahar (15 yaş)	91,25	100	100
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	100	100	100
		Mebrure (60 yaş)	96,25	100	100
		Ortalama	87,91	94,16	97,29

5. Görme Engelliler İçin Metin Düzenleyici

Türkçe Metin Seslendirme sisteminin iki bileşeni de kullanılarak görme engelliler için metin düzenleyici program da geliştirilmiştir. Bu programın asıl hedefi, görme engellilerin diledikleri metinleri yazıp, sesli olarak dinlenebilmesinin sağlanmasıdır. Görme engelli kişi, tuştakımını kullandıkça, yazdığı herşey sesli olarak kendisine dinletilmesi tasarlanmış ve geliştirilmiştir. Görme engelliler için metin düzenleyici programının yeteneklerini aşağıdaki maddelerle özetleyebiliriz.

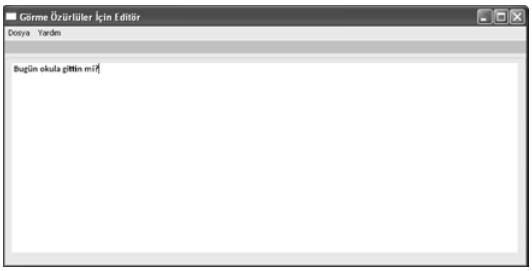
Tuştakımından girilen her harf, sayı veya noktalama işareti kullanıcıya sesli olarak bildirilmektedir.

Harf tuşlanması durumunda seslendirme, sessiz harflerin sonuna "E" seslisi eklenerek, seslilerin ise olduğu haliyle seslendirilmesi sağlanmıştır. Örneğin "MERAK" sözcüğünün yazımı sırasında, tuştakımından tuşlanan "M", "E", "R", "A" ve "K" harfleri sırasıyla "ME", "E", "RE", "A", ve "KE" şeklinde seslendirilmektedir.

Rakamlar tuşlandıkça seslendirilmesi sağlanmıştır. Örneğin tuştakımından girilen "12" sayısı için sırasıyla "BİR" ve "İKİ" seslendirmesi yapılmaktadır.

Noktalama işaretleri tuşlandıkça da tanımlı olduğu şekliyle seslendirilme yapılmaktadır.

Ok tuşları tuşlandığı takdirde imlecin geldiği yerdeki harf, rakam veya noktalama işareti seslendirilmektedir. Örneğin "BUGÜN OKULA GİTTİN Mİ?" cümlesi yazılıyken imleç "?" karakterinin sağında bulunsun (Şekil-10).



Şekil-10: Görme engelliler için metin düzenleyici program

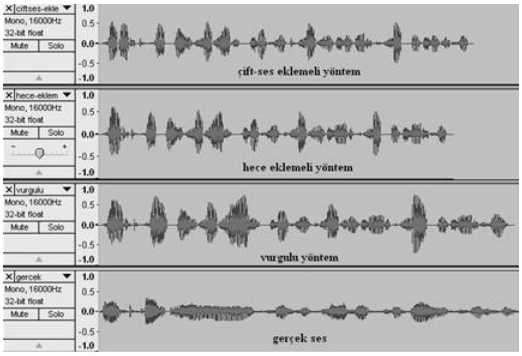
İmleci sola götürmek için ← tuşlandığında imleç "İ" harfi ile "?" arasına gelir ve program "İ" seslendirmesini yapar. Sonra, → tuşlandığında imleç tekrar "?" karakterinin sağına gelir ve bu sefer "SORU İŞARETİ" seslendirmesi yapılır.

- Shift, Backspace, Delete, Page Up, Page Down, Home, End tuşları; normal programlarda olduğu işlevleriyle kullanılmaktadırlar ve bu tuşlar yardımı ile metin üzerinde imleç hareket ettirilebilmektedir. İmlecin yeri değiştiğinde, gelinen yerdeki karakter sesli olarak bilgilendirilmektedir.
- (ALT + S) tuş birlikteliği ile seçili olan metnin, eğer seçili olan metin yoksa tüm metnin seslendirilmesi sağlanmaktadır.
- Metin yazılırken boşluk (SPACE) tuşuna basıldığı takdirde son sözcüğün seslendirilmesi sağlanmaktadır.

6. Sonuçlar ve Öneriler

Bilgisayarla insan ve makine arasındaki sözel iletişim, son yıllarda önemi gittikçe artan bir konudur. Dünyada bu alanda uzun süredir yapılan çalışmalar sonucu, anlaşılırlığı oldukça iyi söz sentezleyiciler geliştirilmiştir. Son yıllarda Türkiye'de de bu alanda yapılan çalışmalar meyvelerini vermeye başlamıştır. Her dilin kendine özgü ses özellikleri mevcut olduğundan, İngilizce söz sentezleyicileri Türkçe söz sentezi için kullanmak mümkün olmamaktadır. İşte bu çalışma, Türkiye'de eksikliği duyulan, anlaşılır Türkçe söz sentezleyiciler konusundaki çalışmalara katkıda bulunabilmek amacıyla yapılmıştır. Çalışma sonucunda, çoklu ortam uygulamaları, konuşma engellilere gerekli iletişim araçlarının temini, görme engellilere okuma araçlarının yapımı gibi konularda kullanılabilecek bir yazılım ortaya çıkmıştır.

Türkçenin söz sentezlemede bilinmesi gereken önemli dilbilgisi kuralları ve sesçil özelliği incelenmiş, birkaç sentez yöntemi tartışılmıştır. Sonuçta, Türkçenin sesçil olması, sondan birçok ek alması ve hece sayısının oldukça fazla olması dikkate alınarak en uygun yöntem belirlenmeye çalışılmıştır. Bu noktada, her bir yöntem için ses üretebilir bir sistemin kurulması ve Şekil-11'deki gibi girilen metnin tüm yöntemler doğrultusunda, ses üretebilmesi sağlanmıştır.



Şekil-11: ""Parkta oynayan çocuklar uçurtma uçurdu" cümlesinin farklı yöntemlerle oluşturulmuş ses dalga şekilleri

Türkçedeki ikisi dış kaynaklı olan sekiz farklı hece tipi üzerinde çeşitli incelemeler yapılmıştır. Her bir hece tipinin ekleme yöntemleri üzerinde formüller gerçeklenmiştir.

Sözcükler arasındaki "ulama", daha heceleme sırasında gerçekleştirilerek, çıkan sesin daha doğal ve anlaşılır olması yolunda olumlu tesir etmiştir.

Çalışmada 65-120 ms uzunlukta 7.845 adet ses dosyası oluşturulmuştur. Bu 16 bitlik kayıtlarda iniş ve çıkışların sert olmaması için algoritma ile tüm dosyalar otomatik olarak elden geçirilmiştir.

Vurgu ve tonlama gibi etkiler, Türkçede anlaşılırlığı önemli derecede değiştiren ses olaylarıdır. Ancak, her yerde geçerli kuralları olmadığı için matematiksel modelini oluşturmak oldukça güçtür. Bu yüzden en çok bilinen özellikleri ile vurgu

çalışması yapılmıştır. Çalışmada asıl ağırlıklı amaç, metinden fonetik seviyede oluşturulan XML dosyası ile XML dosyasından ses üreten iki bileşenin birbirinden tamamen bağımsız çalışabilmesidir. Bu gevşek bağlaşımlı yapı sayesinde, yapılacak doğal dil işleme çalışmaları, bu çalışmaya eklenebilecek ve daha doğal sesler çıkartılabilecektir.

Karşılaşılan bir diğer sorun da, hecelerin, sözcüğün içindeki konumlarına bağlı olarak, seslendirmede değişiklik göstermesinden dolayı tüm heceler için ortak bir yol izlenmiş, bu da vurgu ve tonlamayı olumsuz yönde etkilemiştir.

Türk abecesindeki harflerin, Türkçedeki tüm sesleri karşılamaması da başka bir sorundur. Bu durumlara, özellikle, yabancı kökenli Türkçeleşmiş sözcüklerde rastlanmaktadır. Örneğin, "lale" sözcüğündeki "la" sesi ile "pala" sözcüğündeki "la" sesi birbirinden çok farklıdır. Bunun için öncelikle Türkçe sesçil abecesi çıkartılmıştır ve "lale" gibi sözcüklerde hangi sesin kullanılacağının, aykırı sözcükler sözlüğünden elde edilmesi hedeflenmiştir.

Sonuç olarak, çalışmadan daha doğal sesler elde etmek için frekans alanı üzerinde çalışma yapılmalıdır. Pitch değerleri üzerinde durularak ve doğal dil işleme desteğinin de alınması ile çok doğal sonuçların elde edilebileceği çalışma sonucunda ispatlanmıştır.

7. Kaynakça

[1] Lemmetty S., Review of Speech Synthesis Technology, Helsinki University of Technology, 1999

[2] Allen, J., Hunnicutt, S., Klatt D., From Text to Speech: The MITalk System, Cambridge University Press, 1987

[3] Dutoit T., A Short Introduction to Text-to-Speech Synthesis http://tcts.fpms.ac.be/synthesis/introtts_old.html, alındığı tarih 25.02.2010

[4] http://www.acapela-group.com, alındığı tarih 25.02.2010

[5] Ljungqvist M., Lindström A., Gustafson K., A New System for text-to-Speech and Its Applications to Swedish, ICSLP94 (4) : 1779-1782, 1994

[6] Mönius B., Schroeter J., Santen J., Sproat R., Olive J., Recent Advances Multilingual Text-to-Speech Synthesis, Fortschritte der Akustik, DAGA, 1995

[7] Güldalı K., Türkçe Metin Seslendirme, İstanbul Teknik Üniversitesi, 2009

[8] http://www.nuance.com/realspeak, alındığı tarih 26.02.2010

[9] Festival Project Homepage http://www.cstr.ed.ac.uk/projects/festival, alındığı tarih 26.02.2010

[10] Dutoit T., "An Introduction to Text to Speech Synthesis", pp 26-32, 1997

[11] Wave Dosya Formatı http://ccrma.stanford.edu/courses/422/projects/WaveFormat, alındığı tarih 01.03.2010

[12] SAMPA Türkçe, http://www.phon.ucl.ac.uk/home/sampa/turkish.htm, alındığı tarih 07.03.2010

[13] Türkçe İmla Kılavuzu - Türk Dil Kurumu, 2000

[14] Adalı E., Doğal Dil İşleme, 2010