

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKÇE METİN SESLENDİRME

YÜKSEK LİSANS TEZİ
Tuncay ŞENTÜRK

Anabilim Dalı : Bilgisayar Mühendisliği

Programı : Bilgisayar Mühendisliği

Tez Danışmanı: Prof. Dr. Eşref ADALI

HAZİRAN 2010

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKÇE METİN SESLENDİRME

**YÜKSEK LİSANS TEZİ
Müh. Tuncay ŞENTÜRK
504001576**

**Tezin Enstitüye Verildiği Tarih : 26 Mayıs 2010
Tezin Savunulduğu Tarih : 5 Haziran 2010**

**Tez Danışmanı : Prof. Dr. Eşref ADALI (İTÜ)
Diğer Jüri Üyeleri : Prof. Dr. Coşkun SÖNMEZ (YTÜ)
Yrd. Doç. Dr. Berk ÜSTÜNDAĞ (İTÜ)**

HAZİRAN 2010

Eşime ve oğlum Batu'ya,

ÖNSÖZ

Öğrenim hayatımda ve tez çalışmam boyunca, bana her zaman destek olan ve vakit ayıran tez danışmanım Prof. Dr. Eşref Adalı'ya sonsuz teşekkürlerimi sunarım.

Tez çalışmam süresince bana inanan ve destek olan eşim Pınar ŞENTÜRK'e, sürekli beni teşvik eden Ali Sadık Kumralı'ya, istatistiksel çalışmamda yardımcı olan üç yaşındaki oğlum Batu Şentürk, Arda Göreci, Mebrure Alan, Tarık Özay ve Bahar Deniz'e çok teşekkür ederim.

Haziran 2010

Müh. Tuncay ŞENTÜRK

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	v
İÇİNDEKİLER	vii
KISALTMALAR	ix
ÇİZELGE LİSTESİ.....	xi
ŞEKİL LİSTESİ.....	xiii
ÖZET.....	xv
SUMMARY	xvii
1. GİRİŞ	1
1.1 Çalışmanın Amacı	2
1.2 Literatür Özeti	2
1.3 Önemli Metinden Ses Üretme Sistemleri.....	4
1.3.1 MITalk	5
1.3.2 Infovox	5
1.3.3 Bell Labs TTS	5
1.3.4 SoftVoice	6
1.3.5 ETI Eloquence.....	6
1.3.6 Festival	7
1.3.7 CNET PSOLA.....	7
1.3.8 MBROLA.....	7
1.3.9 Whistler	8
1.3.10 GVZ	8
2. SESBİLİM VE SES ÜRETİMİ.....	9
2.1 Amaç	9
2.2 Sesbilim	9
2.3 Türkçenin Ses Kuralları.....	11
2.4 Türkçe Sözcüklerde Vurgu.....	12
2.5 İnsanda Ses Üretimi	13
2.6 Metinden Söze Geçme Teknikleri.....	14
2.6.1 Mafsallı (articulatory) sentezleme yöntemi	15
2.6.2 Biçimlendirici (formant) sentezleme yöntemi	15
2.6.3 Eklemeli (concatenative) sentezleme yöntemi.....	17
3. TÜRKÇE İÇİN METİN SESLENDİRME SİSTEMİ VE GERÇEKLENMESİ	21
3.1 Amaç	21
3.2 Tasarım.....	21
3.2.1 Metin çözümlemesi aşamasında karşılaşılan problemler.....	21
3.2.1.1 Metin önışleme	21
3.2.1.2 Dilbilimsel analiz	22
3.2.1.3 Ölçü analizi	23
3.2.2 Ses sentezi aşamasında karşılaşılan problemler.....	23
3.3 Türkçe Metin Seslendirme Sistemi	23

3.3.1 Ses dosyaları.....	24
3.3.2 Ses dosyalarının (veritabanı) hazırlanması	27
3.3.2.1 Türkçe sesçil (fonetik) alfabesi	27
3.3.2.2 Türkçede kullanılan hece tipleri	29
3.3.2.3 Mbrola kullanılarak ses veritabanının otomatik olarak oluşturulması.....	32
3.3.2.4 Oluşturulan seslerin genliklerinin dengelenmesi	33
3.3.3 Türkçe Metin Seslendirme Sisteminin Gerçeklenmesi	35
3.3.3.1 Metin işleme bileşeni	35
3.3.3.2 Ses işleme bileşeni	40
Çift-ses (diphone) eklemeli yöntem	41
Hece eklemeli yöntem	42
Farklı hece uzunlukları ile eklemeli yöntem	43
Farklı genlik değerleri ile eklemeli yöntem	44
XML'den okuma	45
Yöntemlerin karşılaştırılması	46
3.3.4 Gerçekleme ortamı	47
3.3.5 Yazılımın diğer yetenekleri	49
3.3.6 Yazılımın çalışabilmesi için gerekli donanım	51
4. TÜRKÇE METİN SESLENDİRME UYGULAMASININ	
DEĞERLENDİRİLMESİ.....	53
4.1 Anlaşılabilirliğin hesaplanması	53
4.2 Cümlelerin belirlenmesi	54
4.2.1 Cümlelerin dinleyicilere dinletilmesi	55
4.2.2 Sonuçların değerlendirilmesi	55
4.2.3 Çalışmanın yaptığı katkılar	58
4.2.4 Görme engelliler için metin düzenleyici program.....	60
5. SONUÇ VE ÖNERİLER.....	63
KAYNAKLAR.....	67

KISALTMALAR

API	: Application Programming Interface
DLL	: Dynamic Link Library
FD-PSOLA	: Frequency Domain-Pitch Synchronous Overlap Add
GUI	: Graphical User Interface
IDE	: Integrated Development Environment
KHz	: Kilo Hertz
LP	: Linear Predictive
MB	: Mega Byte
MOS	: Mean Opinion Score
MS	: Milisaniye
PSOLA	: Pitch Synchronous Overlap Add
TD-PSOLA	: Time Domain-Pitch Synchronous Overlap Add
TDK	: Türk Dil Kurumu
XML	: Extensible Markup Language

ÇİZELGE LİSTESİ

Sayfa

Çizelge 1.1 : Önemli Metinden Ses Üretme Sistemleri	5
Çizelge 1.2 : MBROLA uygulaması için girdi örneği.....	8
Çizelge 2.1 : Harflerin farklı seslerin yerine geçebilmesi	10
Çizelge 2.2 : Türkçede sesli harflerin sınıflandırılması.....	10
Çizelge 2.3 : Türkçede sessiz harflerin sınıflandırılması.....	11
Çizelge 2.4 : İki sessiz ile başlayan sözcüklerde araya gelecek olan ünlü	11
Çizelge 2.5 : Sözcük birimlerinin seçimi.....	19
Çizelge 3.1 : RIFF veri bölgesi (chunk) – 12 sekizli	24
Çizelge 3.2 : FORMAT veri bölgesi (chunk) – 24 sekizli.....	25
Çizelge 3.3 : DATA veri bölgesi (chunk).....	25
Çizelge 3.4 : Türkçe sesçil alfabesi	28
Çizelge 3.4 (devam) : Türkçe sesçil alfabesi	29
Çizelge 3.5 : Hece sonunda bulunabilen iki sessiz	30
Çizelge 3.6 : Hece başında bulunabilen iki sessizler	31
Çizelge 3.7 : Türkçe için oluşturulması gereken ses dosyaları toplamı	32
Çizelge 3.8 : Hecelerin uygun çift-seslere bölünmesi	41
Çizelge 4.1 : MOS seviyeleri.....	53
Çizelge 4.2 : Dinleyicilere dinletilen 10 cümle	55
Çizelge 4.3 : Dinleyici cevaplarından oluşturulan not dağılımı	56
Çizelge 4.4 : Genel ortalamalar	57
Çizelge 4.5 : Dinleyici cevaplarına göre oluşturulan başarı oranları.....	59
Çizelge 4.6 : Noktalama işaretlerinin seslendirilmesi	60

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1 : Metinden ses elde etme aşamaları	1
Şekil 1.2 : Kratzenstein'in ses biçimlendirme cihazları	3
Şekil 1.3 : Wheatstone'un geliştirdiği Kempelen ses cihazı.....	3
Şekil 1.4 : TTS'de önemli kilometre taşları.....	4
Şekil 1.5 : Bell Labs TTS'de İngilizce için kullanılan modüller	6
Şekil 2.1 : İnsanda ses üretme organları	14
Şekil 2.2 : “i” sesine ait formant değerleri.....	16
Şekil 2.3 : Kaskat biçimlendiricinin temel yapısı.....	16
Şekil 2.4 : Paralel biçimlendiricinin temel yapısı.....	17
Şekil 3.1 : Ses sinyalinin örneklenmesi	26
Şekil 3.2 : Ötümlü ve ötümsüz ses.....	26
Şekil 3.3 : Periyodik ses sinyali ve “pitch” periyodu	26
Şekil 3.4 : Mbrola kullanarak ses veritabanının otomatik olarak oluşturulması	33
Şekil 3.5 : Oluşturulan ses dosyalarını içeren veritabanının uygulamaya tanıtılması	34
Şekil 3.6 : Türkçe Metin Seslendirme sisteminin temel akışı	36
Şekil 3.7 : Hece çıkarma ağacı	37
Şekil 3.8 : Metin işleme birimince oluşturulan XML örneği.....	39
Şekil 3.9 : Hecelere parçalama sonrasında hecelerin gösterimi	38
Şekil 3.10 : Türkçe Metin Seslendirme uygulaması ekran görüntüsü	40
Şekil 3.11 : Uygulamada çift-ses eklemeli yöntem seçeneği	42
Şekil 3.12 : Uygulamada hece eklemeli yöntem seçeneği.....	42
Şekil 3.13 : Sesler için süre bilgisinin XML belgesinde gösterimi	43
Şekil 3.14 : Uygulamada değişik ses uzunluklu hece eklemeli yöntem seçeneği	43
Şekil 3.15 : Vurgu katsayısının XML belgesinde gösterimi	44
Şekil 3.16 : “ilk” hecesinin normal ve 1.4 katsayısı ile genliğinin arttırılmış hali....	45
Şekil 3.17 : Uygulamada farklı genlik değerleri ile eklemeli yöntem seçeneği	45
Şekil 3.18 : Uygulamada XML'den okuma seçeneği.....	46
Şekil 3.19 : “halk” kelimesinin tüm yöntemlerdeki ses dalga şekilleri	47
Şekil 3.20 : Dosya menüsü adımları	50
Şekil 3.21 : Düzenle menüsü adımları.....	51
Şekil 4.1 : Görme engelliler için metin düzenleyici programı.....	61
Şekil 5.1 : “Parkta oynayan çocuklar uçurtma uçurdu” ses dalga şekilleri	64

TÜRKÇE METİN SESLENDİRME

ÖZET

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 anlaşılabilirlik istatistiksel olarak ölçülmüştür.

İlk olarak, “çift-ses (diphone) eklemeli yöntem” uygulanmıştır. Anlaşılabilirliği 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. Anlaşılabilirlik 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 çalışmalar için gerekli ses dosyalarının hazırlanması amacıyla önce Türk Dil Kurumunun ses veritabanı kullanılmıştır. Ancak bu veritabanında kelimelerin vurgulu ve iki farklı kişi (erkek ve kadın) tarafından karışık olarak okunmuş olması dolayısıyla çok olumlu sonuçlar elde edilememiştir. 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 anlaşılabilirlik arttırılmıştır. Son olarak bu hecelerin birleşme noktalarında seslerin türlerine göre belirlenen kurallar uygulanarak, gerçek ses dosyalarındaki dalga şekillerine benzer doğallık oluşturulmaya çalışılmıştır.

Hazırlanan program üç ana bileşenden oluşmaktadır:

- Metinden XML dosyası oluşturma : İlk bileşen, girilen metni dilbilgisi kuralları çerçevesinde, belirlenen biçimde bir XML yapısına dönüştürür.
- XML’den ses üretme : Bu bileşen, belirlenen kurallar doğrultusunda hazırlanmış XML dosyasını veya katarını, Türkçe ses dosyasına dönüştürür.
- Kullanıcı arayüzü : Programın kullanılabilmesi için hazırlanmış arayüz bileşenidir. Her iki bileşen, birbirine bağlanmıştır ve görsel arayüz ile kullanıcının girmiş olduğu metin, yine kullanıcının belirlemiş olduğu yöntem ile ses dosyasına dönüştürülüp, seslendirilir.

Tüm yöntemlerin ayrı ayrı anlaşılabilirliğinin tespit edilebilmesi için; cümleler, farklı yaş gruplarındaki insanlara dinletilmiş ve alınan cevaplara göre belirli formül yardımı ile yüz üzerinden puan verilecek şekilde hesaplama yapılarak, bir matriste sunulmuştur.

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

TURKISH TEXT TO SPEECH SYNTHESIZER

SUMMARY

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 statistically 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.

The Turkish Language Association's (TDK) database is used to prepare the necessary audio files in the begining of this study. However, in this database the sound of words were accented, and the database was vocalized by two different people (men and women) therefore favorable results could not be achieved. 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. Finally, more natural sounds which have a wave shape similar to real audio files were created by applying the rules, determined according to the type of sound, to the syllables vanishing point.

The program consists of three main components:

- Text to XML: the first component converts the text to the specified XML format by given grammar rules framework.
- XML to sound: This component converts the XML file or string, which has been prepared in accordance with the rules specified, to Turkish audio files.

- Graphical User interface: the interface is the component designed to use the program. Both components are linked together. The text entered by the user interface, is converted to audio file utilizing the method selected by the user, and then vocalized.

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.

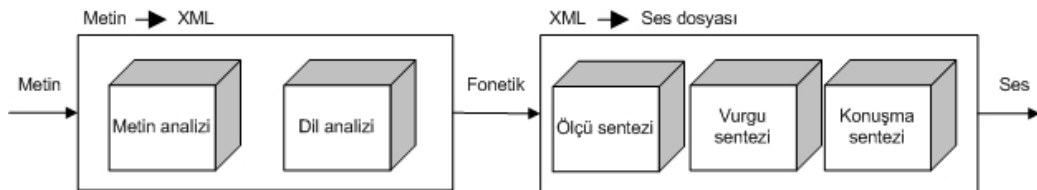
Finally, a text editing software program is developed to help the visually impaired edit text without the need for a screen image.

1. GİRİŞ

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. Etkileşimli dizgelerde makina-insan arasındaki iletişimin sözel olması, yani bilginin kaynağının doğrudan kulağı uyarması, bilgi gelişini kullanıcının sürekli izlemesi zorunluluğunu ortadan kaldırır[1]. Dolayısıyla, uyarıları sözle ileten araçların kullanımının artması, vakit alıcı birçok denetimi azaltır. Bilginin kullanıcıya sözel olarak iletilmesine olanak sağlayan söz sentezleyicilerinde metin, girdi; ses, çıktı olarak kullanılmaktadır. Makina-insan haberleşmesinin, konuşma engellilere gerekli iletişim araçlarının temini, görme engellilere okuma aletlerinin yapımı, sesli bilgilendirme sistemleri, lisan eğitim programları gibi yaygın kullanım alanları bulunmaktadır.

Metinden ses elde etme yöntemine karakteristik olarak üç yöntemle değer biçilir: doğruluk (metnin doğru bir şekilde değerlendirilmesi, kısaltmalar, sayılar, e-posta adresleri, vb.), anlaşılabilirlik (yapılan sınama ile oluşturulan sesin anlaşılabilirliğinin yüzde cinsinden tespit edilmesi) ve doğallık (oluşturulan ses sinyallerinin doğal ses sinyalleri ile benzeşmesi)[2].

Metinden ses üretme, Şekil 1.1’de gösterildiği gibi iki temel bileşenden oluşur. Her iki bileşenin birbirine bağımlılığı en az seviyeye indirgenmiştir (İngilizce terminolojide iki bileşenin birbirine bağımlılığının en aza indirgenmesine “*loosely coupled*” denmektedir). Bu sayede iki bileşen farklı uygulamalarda bağımsız şekilde kullanılabilir şekilde tasarlanmıştır. İlk bileşen metnin, dilbilimsel kurallar çerçevesinde, ses sinyallerine dönüştürülmek üzere belirlenecek bir biçime dönüştürülmesini sağlamaktadır. Bu çalışmada, bunun için XML kullanılmıştır.



Şekil 1.1 : Metinden ses elde etme aşamaları

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.

Çalışmada ayrıca, kullanıcının bu iki bileşenin detayını bilmesini gerektirmeden, sadece girdiği metni seslendirebilmesi amacıyla, basit bir arayüz de hazırlanmıştır.

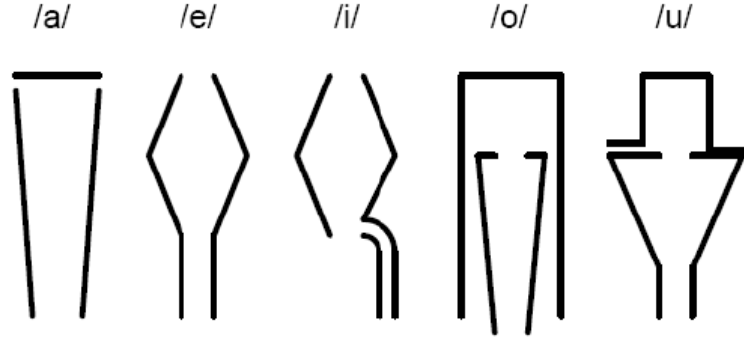
Son olarak, tüm bu çalışmaların da kullanıldığı bir arayüz ile görme engelliler için metin düzenleme programı hazırlanmıştır. Bu program ile bilgisayar ekranına bakmadan, sadece tuştakımı ile komutların ve yazılan metnin seslendirilmesi sağlanabilmektedir.

1.1 Çalışmanın Amacı

Bu çalışmanın amacı Türkçe'nin yapısal özelliklerini kullanarak, metinden söze geçme (text-to-speech) olanağı veren metin seslendiricisi gerçekleştirmektir. Bu amaçla Türkçe'nin sesçil özellikleri, değişik sesler arasındaki ilişkiler araştırılmış ve bu özelliklerden yararlanılarak en doğal şekilde sentezlenip, mümkün olduğunca doğal bir sonuç verecek şekilde ses birimleri oluşturulmuştur. Türkçe metin seslendirme sistemini gerçekleştirirken, daha önceden yapılmış ve kabul görmüş ses birleştirme yöntemlerinin geliştirilmesi, seçime bağlı olarak uygulanabilmesi ve sonucunda hangi yöntemin daha uygun olduğunun deneysel olarak ortaya çıkarılması hedeflenmektedir. Ayrıca, görme engellilerin kullanımı amacıyla, bilgisayar ekranı gerektirmeden; metin yazma ve düzenleme çoğul ortamının hazırlanarak kullanıma sunumu amaçlanmaktadır.

1.2 Literatür Özeti

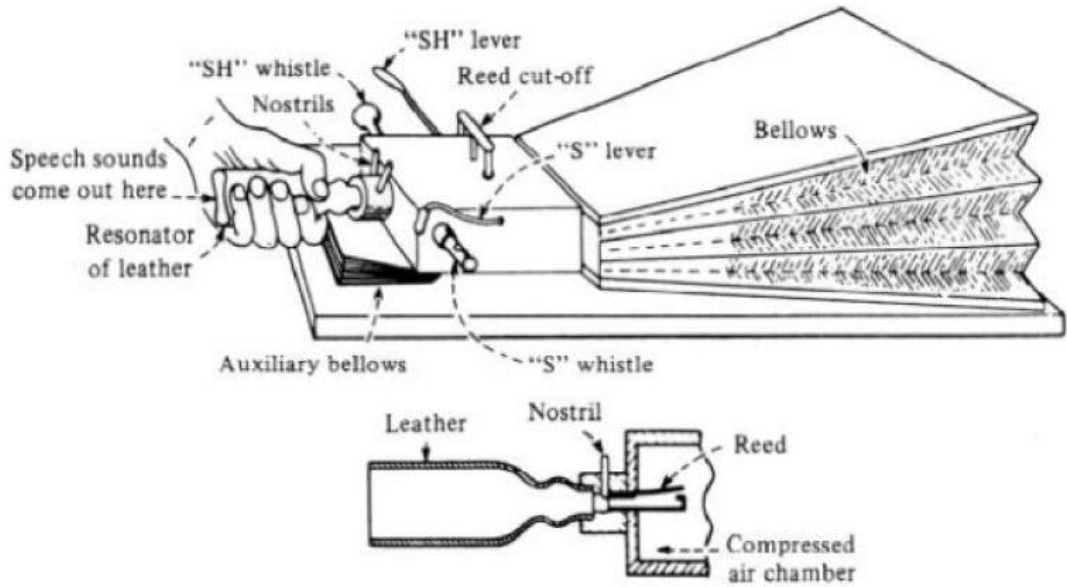
Ses sentezleme sistemleri ilk bilgisayarın yapımından hemen sonra başlamıştır. 1779 yılında Rus professor Christian Kratzenstein, beş sesli harfin (/a/, /e/, /i/, /o/, ve /u/) fiziksel farklılıklarını incelemiş ve St. Petersburg'da bu sesleri üreten bir sistem yapmıştır[3, 4]. Bu sistemin temel ses çıkaran aletleri Şekil 1.2'de görülmektedir[5].



Şekil 1.2 : Kratzenstein'in ses biçimlendirme cihazları

1791 yılında Viyana'da, Wolfgang von Kempelen "Akustik Mekanik Konuşma Makinesi" adlı bir makine geliştirmiş, bazı ses ve ses kombinasyonlarını bu makine vasıtasıyla çıkarmayı kısmen başarmıştır[6, 7].

1800'lü yılların ortalarında Charles Wheatstone, Kempelen'in cihazını geliştirmiş ve tüm seslilerin yanında birçok sessiz harfi de seslendirir hale getirmiştir[8]. Birçok yönden olumlu sonuç almış, heceleri hatta bazı kelimeleri dahi bu cihaz (Şekil 1.3) [5, 8] ile üretebilir hale getirmiştir.



Şekil 1.3 : Wheatstone'un geliştirdiği Kempelen ses cihazı

1838 Willis, ses yoluna benzeyen borularla bazı sesleri elde etmeyi başaramamıştır. Ayrıca seslinin ses kalitesinin, kullanılan borunun çapına değil, sadece uzunluğuna bağlı olduğunu bulmuştur[4, 9].

1922’ de Stewart, ilk elektrikli ses sentez cihazını tanıtmıştır. Bu cihaz sadece seslileri çıkarabilmekteydi[7, 9]. Ancak literatürde, Homer Dudley tarafından 1939 yılında yapılan VODER (Voice Operating Demonstrator) ilk olarak kabul edilmektedir[5, 9, 10]. Bu çalışmalar, oluşturulan ses kalitesi pek de beklenen düzeyde olmamasına rağmen, yapay ses oluşturma açısından önemli adım olarak nitelendirilmiştir.

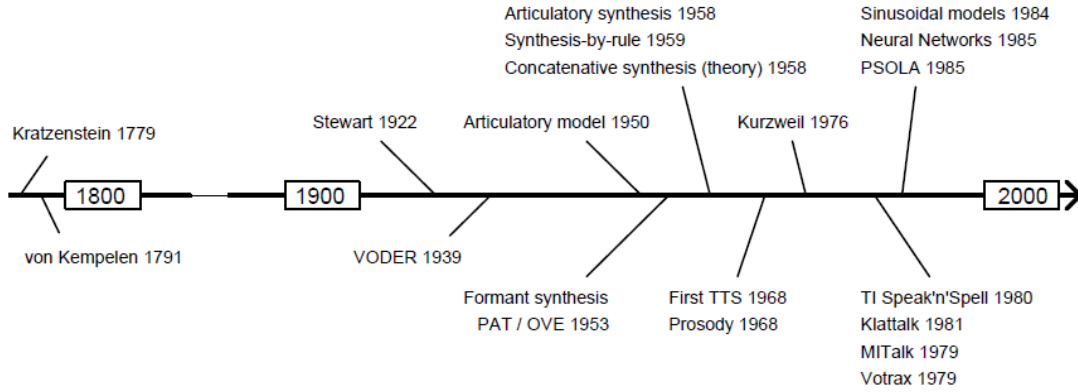
1953’ de ilk biçimlendirici (formant) sentezleyici PAT (Parametric Artificial Talker), Walter Lawrence tarafından tanıtıldı[5, 9].

1958’ de ilk mafsallı (articulatory) sentezleyici, DAVO (Dynamic Analog of VOcal Tract), George Rosen tarafından MIT’ de geliştirilmiştir[11].

Yakın geçmişte yapılan ve metin sentezleme konusunda önemli rol oynayan çalışmalara bir sonraki başlıkta değinilecektir.

1.3 Önemli Metinden Ses Üretme Sistemleri

Bu bölümde, yakın geçmişte yapılan ve ses getirmiş metinden ses üretme çalışmalarından bir kısmı ele alınacaktır. Ses sentezleme açısından önemli kilometre taşları Şekil 1.4’de tarih sırasında listelenmiştir[20].



Şekil 1.4 : TTS’de önemli kilometre taşları

Çizelge 1.1’de yapılan çalışmalar ve bu çalışmaların dahil olduğu yöntemler listelenmiştir.

Çizelge 1.1 : 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

1.3.1 MITalk

J. Allen, S. Hunnicutt, D. Klatt tarafından 1979 yılında MIT laboratuvarlarında geliştirilmiştir. Biçimlendirici (formant) temellidir. Günümüzde kullanılan teknolojiye ve birçok çalışmada temel teşkil etmiştir[12, 13].

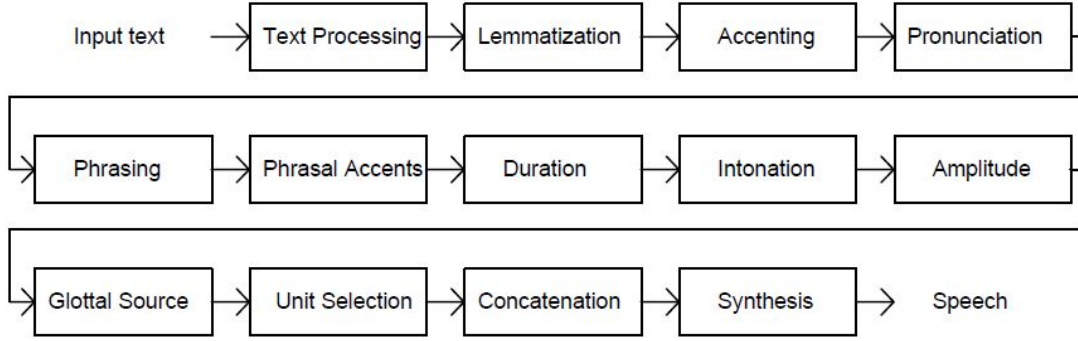
1.3.2 Infovox

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. İngilizce (Amerikan ve İngiliz), Almanca, Fransızca, İspanyolca, İtalyanca, Fince, İsveççe, Danimarkaca, İzlandaca ve Türkçe dahil olmak üzere birçok dil desteği bulunmaktadır. Üretilen ses kalitesi genel olarak anlaşılırdır. Erkek, kadın, çocuk dahil olmak üzere beş farklı ses ile konuşma üretilebilmektedir[14, 15].

1.3.3 Bell Labs TTS

Çift-ses (diphone) ve üçlü ses (triphone) ekleme (concatenative) yöntemine dayanmaktadır ve İspanyolca, İtalyanca, Rusça, Romence, Çince ve Japonca desteği bulunmaktadır[16]. Oldukça modüler bir yapısı olduğu için, farklı grupların aynı proje üzerinde çalışmasını oldukça kolay hale getirmiştir.

Şekil 1.5’de İngilizce için geliştirilmiş olan modüllerin etkileşimi görülmektedir[20]. Bu modüllerden de anlaşılacağı üzere, ticari uygulamanın metinden ses üretme özelliğinin yanında gramer yeteneği de oldukça başarılı olmuş ve desteklediği dillerde olumlu sonuçlar vermiştir[17].



Şekil 1.5 : Bell Labs TTS’de İngilizce için kullanılan modüller

1.3.4 SoftVoice

TTS konusunda 25 yıldan fazla tecrübesi 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. Erkek, kadın, çocuk, robot ve uzaylı gibi ses türleri bulunmaktadır. Diğer ürünlere nazaran çok başarılı sonuçlar üretememiştir ancak kolayca yeni ses eklenebilme özelliği sayesinde, çoğul ortam ürünlerinde en fazla tercih edilen çözüm olmayı başarmıştır[20].

1.3.5 ETI Eloquence

Eloquent Technology (ABD) tarafından geliştirilmiş, eklemeli yöntem kullanan, çoklu dil desteği sunan bir sistemdir. İngiliz ve Amerikan İngilizcesi, Almanca, Fransızca, İtalyanca ve İspanyolca desteği vardır ve her bir dil için kullanıcılar tarafından kolaylıkla uyarlanabilen, yedi farklı ses tonu (erkek, kadın ve çocuk dahil) bulunmaktadır. Örneğin baş genişliği (ses yolu ile doğru orantılı), nefes (yüksek değerde fısıltı gibi konuşma), periyodik ötüm değişikliği (pitch fluctuation, yüksek değerde monoton bir konuşma), ses seviyesi (volume) gibi parametrelerle oluşan sesler uyarlanabilmektedir. 2001 yılında Speechworks firması ile birleşmiş, 2003 yılında Nuance Communications firması tarafından satın alınmıştır. Ürünün bugünkü adı Realspeak’tir[18].

1.3.6 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. Kişisel ve araştırma amaçlı kullanımı için açık kaynak kodlu ve ücretsizdir, ayrıca birçok Linux dağıtımının içinde kişisel bilgisayarlarda kullanılabilmektedir[19].

1.3.7 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 İngilizcesi, Fransızca, İspanyolca ve Almanca desteği bulunmaktadır. Sesler, PSOLA (Pitch Synchronous Overlap Add) algoritması ile normalize edilerek birleştirilmiştir. PSOLA algoritmasının birkaç farklı alt dalı bulunmaktadır. TD-PSOLA (Time Domain Pitch Synchronous Overlap Add) en çok bilinen ve kullanılan yöntem iken, FD-PSOLA (Frequency Domain) ve LP-PSOLA (Linear-Predictive) çeşitleri de bulunmaktadır[20].

1.3.8 MBROLA

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].

İkili ses ekleme yöntemi uygulanmıştır ve girdi olarak metin yerine ses (phoneme), süre ve frekans bilgileri alır ve çıktı olarak 16 kHz frekansında 16 bitlik örnek verileri oluşturur. Bu yüzden tam olarak metin seslendirme sistemi olarak anılmaz. Daha çok metin seslendirme sistemlerinde düşük seviye (low level) sentezleme aracı olarak kullanılır[20]. Çizelge 1.2'de “**an**” hecesinin (başında sessizlik ile birlikte) MBROLA uygulaması için girdi örneği bulunmaktadır. Her satır, sesin SAMPA'daki karşılığı, milisaniye cinsinden uzunluk değeri ve ikililer halinde frekans noktaları gelmektedir. Bu ikililerden ilki o ses içerisinde kaçınıcı yüzdede frekans noktası olduğunu, ikincisi ise Hz cinsinden frekans değerini vermektedir.

Mbrola tanımlanmış az sayıdaki frekans noktaları arasına interpolasyonla düz çizgiler çekerek frekans eğrisini oluşturmaktadır. Çizelge 1.2’de sadece iki frekans noktası tanımlanmış :

- Birincisi a sesi içerisinde $70 * \%10 = 7$ (7. milisaniyede ve değeri 100Hz)
- İkincisi n sesi içerisinde $60 * \%80 = 42$ (42. milisaniyede ve bütün içerisinde $70 + 42 = 112$. milisaniyede ve değeri 120 Hz.)

Çizelge 1.2 : MBROLA uygulaması için girdi örneği

a 70 10 100
n 60 80 120

Mbrola’nın Türkçe dahil olmak üzere birçok dil için erkek ve kadın sesi olmak üzere desteği bulunmaktadır.

1.3.9 Whistler

Microsoft Whistler (**W**hisper **H**ighly **I**ntelligent **S**tochastic **TaLkER**), Amerika’da Microsoft araştırma laboratuvarlarında (ABD) geliştirilen ve eğitilebilir amaçlı bir çalışmadır. Ses üretiminde eklemeli yöntem kullanılmış, eğitilebilir olma modülünde ise Hidden Markov Model (HMM) esas alınmıştır[20].

1.3.10 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.

2. SESBİLİM VE SES ÜRETİMİ

2.1 Amaç

Her dilin kendine özgü farklı dilbilgisi kuralları vardır ve yapılacak sentez, o dilin dilbilgisi özelliklerine uygun olmalıdır. Bu bağlamda sentez ve dilbilgisi analizi kavramları oldukça kapsamlı konular içermektedir.

Bu bölümde, bu terminolojinin daha iyi anlaşılabilmesi için; sesbilimi, dilbilgisi kuralları ve insanda ses üretimi gibi temel konulara değinilecektir.

2.2 Sesbilim

Dilin en temel bileşeni ses ve bu seslere karşılık düşen harflerdir. Harf, yazılı olup göze hitap eder. Ses ise, aksine, sözlüdür ve kulağa hitap eder. Sesler, söylenilen ve işitilen, yani konuşulan küçük dil birliktelikleridir. Seslerin yazıya çevrilmesinde kullanılan işaretlere “harf” denir[24].

Bir dildeki sesleri karşılayan harflerin hepsinin birden meydana getirdiği, belirli sırası olan topluluğa “alfabe” denir[25]. Türkçe alfabesinde 29 harf vardır ve en genel olarak sesliler (ünlüler) ve sessizler (ünsüzler) olmak üzere sınıflandırılmışlardır.

a, b, c, ç, d, e, f, g, ğ, h, ı, i, j, k, l, m, n, o, ö, p, r, s, ş, t, u, ü, v, y, z

Yazı dilindeki 29 harfe karşılık, konuşma sırasında kullanılan fazladan seslerle birlikte 36 adet ses olduğu söylenebilir. Örneğin a, e, g, k ve l ile birlikte Çizelge 2.1 :‘de gösterildiği üzere 10 tane ses karşılanmaya çalışılır.

Çizelge 2.1 : Harflerin farklı seslerin yerine geçebilmesi

Ses	Örnek Kelime(ler)
a	zar
a	saat, alim
e	yemek
e	el, vermek
g	geri, gemi
g	gaga, gayda
k	kalın, katı
k	keser, kivi
l	sal, kalın
l	lastik, lazım

“Sesliler (ünlüler)”, ciğerden gelen hava akımının, ses yolunda ve ağızda bir engel ile karşılaşmadan çıkardığı seslerdir. Türkçede ünlüler dilin, alt çenenin ve dudakların aldıkları durumlara göre Çizelge 2.2’de gösterildiği gibi sınıflandırılır[23].

- Dilin aldığı duruma göre : kalın – ince
- Alt çenenin aldığı duruma göre : dar – geniş
- Dudakların aldığı duruma göre : yuvarlak – düz

Çizelge 2.2 : Türkçede sesli harflerin sınıflandırılması

	Düz		Yuvarlak	
	Geniş	Dar	Geniş	Dar
Kalın	a	ı	o	u
İnce	e	i	ö	ü

Sözcüklerin birinci hecelerinde sekiz ünlü sesbirimden herhangi biri bulunabildiği halde, yapısal uyum nedeniyle sözcüklerin ikinci hecesinde bulunabilecek ünlü sayısı dörde inmektedir. Eklerde ise, “yor” dışında, yalnızca iki ünlü sesbirim bulunabilmektedir[1].

Ses yolundan değişik biçimli temas, engel veya takıntılara uğrayarak oluşan sesler “sessizler (ünsüzler)” olarak adlandırılır. Sessiz harfler 21 tanedir ve Çizelge 2.3’te olduğu gibi sınıflandırılırlar.

- Ses kırılganlığının titreşimine göre : sert – yumuşak
- Ses yolunun tıkalı ya da açık oluşuna göre : sürekli – süreksiz

- Sesin çıktığı yere göre : dudak – damak – diş – gırtlak

Çizelge 2.3 : Türkçede sessiz harflerin sınıflandırılması

	Sert		Yumuşak	
	Sürekli	Süreksiz	Sürekli	Süreksiz
Dudak	f	p	m, v	b
Diş	s, ş	ç, t	j, l, n, r, z	c, d
Damak		k	ğ, y	g
Gırtlak	h			

2.3 Türkçenin Ses Kuralları

Türkçede tüm kurallar, seslere ilişkin organların hareketlerine uygun yapıdadır, herhangi bir zorlanma durumunda, sözcük içindeki sesler değişime uğrar. Sessizlerin yanyana gelmesi, eklerin ulanması ve yabancı sözcükleri uyarlamada bu kurallara uyulmaktadır. Bu kurallar metin analizinde ve söz sentezinde önemli rol oynamaktadır.

- Türkçe sözcüklerin başında iki ünsüz art arda gelmez, oysa dış kaynaklı bazı sözcüklerin “yazılışlarında” sözcük başında iki ünsüzün yan yana geldiği görülmektedir. Örneğin standart sözcüğü, konuşma dilinde, araya uygun bir ünlü getirilerek “sı^ındart” şeklinde seslendirilir. İki sessiz arasına gelecek ünlü Çizelge 2.4’teki kurala göre belirlenebilir.

Çizelge 2.4 : İki sessiz ile başlayan sözcüklerde araya gelecek olan ünlü

Hecede bulunan sesli	İki sessiz arasına gelecek sesli	örnek
a, ı, o	ı	kıral tırole ı büs
u	u	guru u p
ü	ü	bür ü t
e, i, ö	i	tire i n tiri i lyon

- Kökeni Türkçe olan kelimelerde uzun ünlü yoktur. Uzun ünlü, Arapça ve Farsçadan Türkçeye giren kelimelerde görülür: şair (şa:ir), numune (numu:ne), iman (i:man), adalet (ada:let), badem (ba:dem), beraber (bera:ber), idare (ida:re), ifade (ifa:de), isaret (isa:ret), kaide (ka:ide), rica (rica:), şive (şi:ve), şube (şu:be), vali (va:li), vefa (vefa:). Bu örneklerde iki noktadan önceki harfin gösterdiği ses uzun ünlüdür ve uzun söylenir[24].

- Uzun ünlülü kapalı hecelerle biten kelimeler ünlüyle başlayan ek aldıklarında veya yardımcı fiillerle birlikte kullanıldıklarında, açık hale gelen hecenin ünlüsündeki uzunluk çoğunlukla yeniden ortaya çıkar: esas / esasen (esa:sen); hayat / hayatı (haya:tı); kanun / kanunen (ka:nu:nen); ruh / ruhum (ru:hum); usul / usulü (usu:lü); vicdan / vicdanen (vicda:nen)[24].
- Arapçadan alınmış bazı kelimelerde, gırtlak ünsüzü hecenin sonunda yer almaktadır. Bu tür kelimelerde gırtlak ünsüzü Türkçe söyleyişten tamamen kalkmakta ve kendisinden önceki ünlünün uzamasına yol açmaktadır: dava (da:va), mana (ma:na), memur (me:mur), tecil (te:cil), telif (te:lif), tesir (te:sir)[24].
- Konuşmada art arda gelen kelimelerden, birincisinin sonundaki ünsüzün, ikincisinin başındaki ünlüye ses bakımından bağlanarak söylenmesine “ulama” denir[24]. Örneğin “satın almak” ifadesi hecelere yazım olarak “sa-tın al-mak” şeklinde ayrılırken, seslendirmede “sa-tı-nal-mak” şeklinde bir birliktelik söz konusudur.

2.4 Türkçe Sözcüklerde Vurgu

Sözcüklerin söylenişinde her hecenin üzerine aynı kuvvetle basılmaz. Konuşma sırasında sözcük ve cümlelerin tekdüze heceler dizisi halinde ortaya çıkmamasının nedeni budur. Sözcüklerde, kuvvetli söylenen hece üzerindeki baskıya “vurgu” denir. Türkçe, yumuşak vurgulu, hafif dalgalı bir dil olup, seslendirmede titiz davranmak gereklidir. Sözcüklerin değişik vurgularda söylenmesi, anlam farklılıkları doğurmaktadır. Örneğin yükselen vurgu “ya” ve düşen vurgu “ya” sözcüklerinde aynı “ya” eki farklı anlamlar taşır. İlki “gördün mü?”, ikincisi “öyle mi?” anlamlarına gelmektedir[1]. Türkçe sözcüklerde vurgu aşağıdaki yöntemlere göre bulunabilir[25] :

- Türkçede vurgu genel olarak ya ilk ya da son hecededir. Vurgu son hecede ise başta, ilk hecede ise sonda ikinci vurgu bulunur.
- Orta hece(ler) vurgusuzdur.
- Genel olarak sözcüklerin büyük bir kısmında vurgu, son hecededir.

- Yer adları ve coğrafi isimlerde vurgu ilk hecede bulunur.
- Sonu –ya ile biten yer adlarında vurgu sondan bir önceki hecede bulunur.
- Türkçede henüz tam benimsenmemiş yabancı kökenli sözcüklerde vurgu ilk hecededir (banka, posta, radyo).
- Sözcük ne olursa olsun, bütün hitaplarda vurgu ilk heceye geçer.

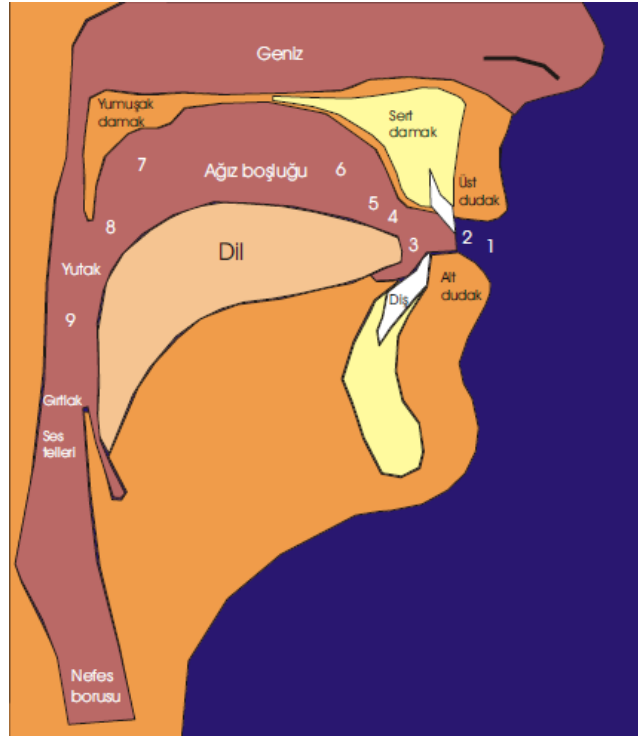
2.5 İnsanda Ses Üretimi

İnsan sesi, ses organlarının istemli hareketleri sonucunda üretilmektedir. Dolayısıyla öncelikli olarak insan ses üretme yöntemlerinin incelenmesi gerekmektedir. Şekil 2.1’de insanda ses üretme organları gösterilmektedir[23]. İnsanda ses üretimi için ana enerji kaynağı diyafram ve ciğerdir. Ciğerlerimizden gelen hava, nefes borusu içinden, gırtlaktaki ses tellerinden geçerek, yutak ve burun boşluğuna ulaşır ve ağız ve burundan dışarı ses olarak çıkar. Ses sistemimizdeki en önemli bölge nefes borusunun bitiminde, ses tellerinin arasında bulunan V şeklindeki açıklıktır. Burada hava akımı ayarlanarak sesli ve sessiz harfler oluşturulur. Ses üretiminde dil, dudak ve dişlerin etkisi de bulunmaktadır. Ayrıca çene; ağız ve geniz boşluğunu değiştirdiği için ses üretim organı olarak sayılmaktadır. Ses tellerinin temel titreşim frekansı erkeklerde 110 Hz, kadınlarda 200Hz ve çocuklarda 300 Hz civarındadır. Ağız boşluğunun boyutu ve yapısı damak, dil, dudaklar, yanaklar ve dişlerin hareketi ile değişir. Burun boşluğunun boyutu ise sabittir ve buraya giren hava akımı yumuşak damak ile kontrol edilir[27,28]. Şekil 2.1’de seslerin üretim bölgeleri gösterilmektedir.

- 1) Dudak sesleri (*bilabial*) : Dudakların şekline göre çıkan seslerdir (b, p, m)
- 2) Dudak – diş sesleri (*labiodentals*) : Alt dudak ve üst dişler arası sesler (f,v)
- 3) Diş sesleri (*dental*) : Dil ucu ve dışardaki sesleri (d, t)
- 4) Diş yuvası (*alveolar*): Dil ucu ile diş eti arası (n, r, s, z)
- 5) Geriye dönük (*retroflex*) : Dil, geriye dönük bir hal alır (r)
- 6) Ön damak sesleri (*palatal*) : Dil ortası ile sert damak arası (ç, c, j, ş, y)
- 7) Damaksız sesler (*velar*) : Dilin gerisi ile damak arası (k, g)

8) Küçük dil sesleri (*uvular*): Dilin gerisi ile küçük dil arası (q)

9) Gırtlak sesleri (*glottal*) : Gırtlak ile nefes borusu arası (h)[23]



Şekil 2.1 : İnsanda ses üretme organları

2.6 Metinden Söze Geçme Teknikleri

Yapılan araştırmalar sonucunda, bir metni seslendirmek için temel olarak üç farklı sentezleyici yöntemi geliştirilmiştir.

1. Mafsallı (articulatory) sentezleyiciler, insanın ses üretme sistemini modellemeyi amaçlar.
2. Biçimlendirici (formant) sentezleyiciler, temel frekansın modellenmesi ve filtre aracılığıyla şekillendirilerek istenilen sesin üretimini amaçlar.
3. Eklemeli (concatenative) sentezleyiciler, önceden kaydedilmiş, değişik uzunlukta doğal seslerin birleştirilmesi ile ses üretimini amaçlar.

Günümüzde biçimlendirici (formant) ve eklemeli (concatenative) sentezleyiciler daha çok kullanılmakta ve tercih edilmektedir. Biçimlendirici sentezleyiciler uzun yıllar daha popüler iken günümüzde ekleme yöntemi daha öne çıkmaya başlamıştır. Mafsallı sentez ise halen oldukça karmaşık ve gerçekleştirmesi zor yöntem olduğu için pek tercih edilmemektedir. Ancak doğal seslerin üretimi sağlanabilecek teknoloji bulunduğu takdirde tüm yöntemleri hızlı bir şekilde geride bırakabilir[20].

2.6.1 Mafsallı (articulatory) sentezleme yöntemi

Mafsallı (telaffuz) sentezleyiciler, insanda ses üretiminde kullanılan organların, mümkün olduğunca iyi modellenmesini amaçlamaktadır. Bu yüzden ses üretme yöntemleri arasında en karmaşık ama bir o kadar da umut verici olanıdır. Ancak oldukça fazla işlem gücü gerektirdiği için ve zor olduğu için –akademik çalışmalar dışında- yeteri kadar ilgi görememiş ve diğer yöntemlerdeki gibi başarılı sonuçlar elde edilememiştir[29,30].

Mafsallı sentezleme yönteminin temelini, insanın ses üretmesinde etkili olan diyafram, akciğer, ağız boşluğu, dil ve ses tellerinin tamamının oluşturduğu yapının modellenmesi ve bu modelden çeşitli süzgeç katsayılarının elde edilmesi oluşturmaktadır. Çeşitli söz öğeleri için bulunan süzgeç katsayılarının beyaz gürültü ve/veya periyodik darbe katarı gibi girişlere uygulanması ile elde edilen çıkışlarla ses üretme işlemi gerçekleştirilir. Bu katsayılardan; dudak aralığı, dudak çıkıntısı, dil ucu yüksekliği, dil ucu duruşu, dil yüksekliği, dil duruşu ve damak aralığı gibi parametreler kontrol parametresi olarak, gırtlak açıklığı, ses tellerinin gerginliği ve ciğerlerden gelen hava basıncı uyarım parametresi olarak kullanılmaktadır. Konuştuğumuzda ses yoluna ait kasların etkisiyle ses yolu değişikliğe uğrar ve değişik sesleri çıkarmamızı sağlar. Bu yöntemi kullanarak geliştirme yapanlar X ışını altında gerçek konuşma analizi yaparak modelleme için veri toplamaktadırlar. Bu veriler iki boyutlu olduğu için üç boyutlu olan ses yolunun modellenmesi zor olmaktadır. Gelişen işlemci güçleri ve bilgisayarlı üç boyutlu görüntüleme sistemlerinin gelişimi ile gelecekte mafsallı sentezleme yönteminden olumlu sonuçlar alınması beklenmektedir[20].

2.6.2 Biçimlendirici (formant) sentezleme yöntemi

Metnin sese dönüştürülebilmesi için izlenen bir başka yöntem de yapay ses üretme, biçimlendirici (formant) sentezleme yöntemidir ve geçtiğimiz yıllarda oldukça popüler hale gelmiştir. Genel olarak, paralel ve kademeli (cascade) şeklinde iki temel yapı kullanılır. Fakat en iyi performans, iki yapının çeşitli şekillerde birleştirilmesiyle oluşur[32]. Biçimlendirici sentezi ayrıca, sonsuz sayıda ses üretimine imkan sağladığı için, eklemeli yöntemlere göre daha esnek bir yapıya sahiptir. DECTalk, MITalk, Infovox'un eski sürümleri bu yöntemi kullanmışlardır.

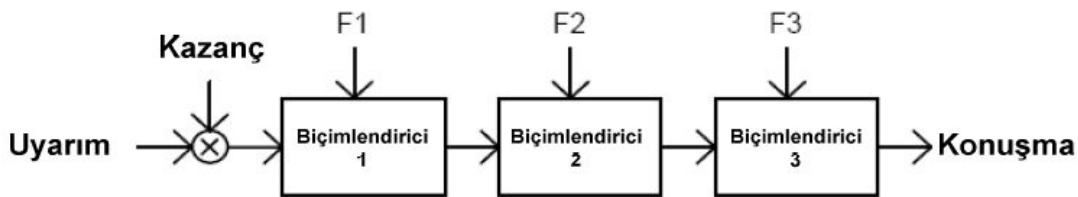
Seslerin tek bir ses frekansı olmadıkları; birleşik ses sinyallerinden oluştuğu görülmüştür. Bir başka deyişle, bir ünlü ses, bir temel frekans bileşeni ve buna ek olarak daha düşük genlikli frekans bileşenlerinden oluşmaktadır. Bir sesi oluşturan bu frekans bileşenlerine "formant" adı verilmektedir[23]. Anlaşılır ses üretebilmek için en az üç formanta ihtiyaç duyulurken, yüksek kalitede konuşma üretebilmek için beş formant gerekmektedir[20]. Şekil 2.2’de IPA tarafından verilen “i” sesine ait formant değerleri görülmektedir.

IPA	Latin	F1	F2	F3
i	IY	270	2290	3010

Şekil 2.2 : “i” sesine ait formant değerleri

Kural tabanlı formant sentezleme yöntemi, adından da anlaşılacağı üzere, formant sentezleme yöntemi kullanarak, istenen telaffuzu üretebilmek için belli başlı parametrelerle karar vermeye dayanmaktadır. Bu parametrelere örnek olarak, temel frekans (F0), formant frekans ve genlik değerleri (F1..F3 ve A1..A3), ses uyarım ve tonlama değeri (V0) verilebilir.

Kademeli (cascade) formant sentezi, birbirine seri bir şekilde bağlanmış, bant geçiren yapıda sesi aksettirici birimlerden oluşur. Bir sesi aksettirici birimin çıkışı, diğerinin girişi şeklindedir. Kademeli yapı, kontrol bilgisi olarak sadece formant frekanslarına ihtiyaç duyar. Kademeli formant sentezinde kullanılan temel yapı Şekil 2.3’de gösterildiği gibidir[20,32].

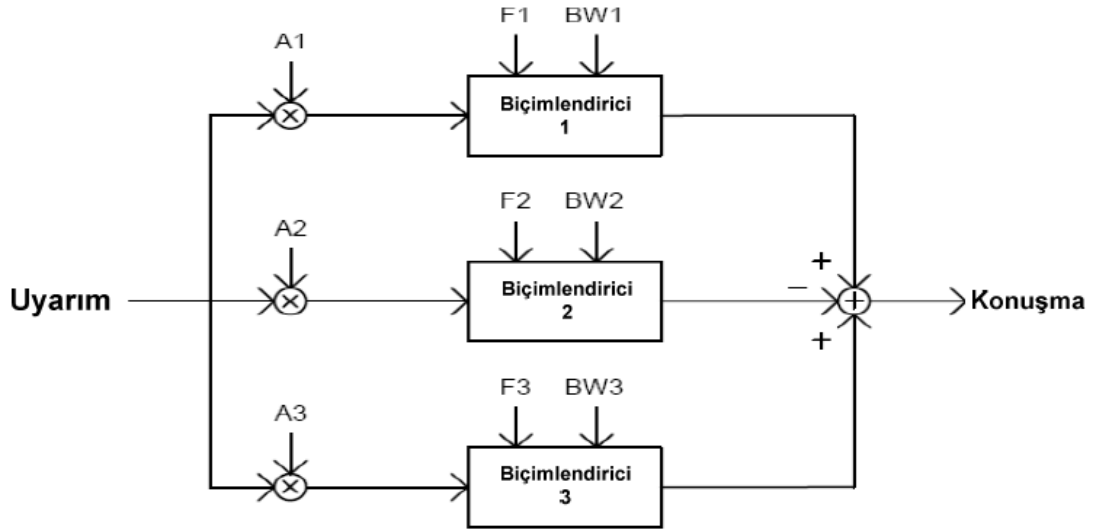


Şekil 2.3 : Kaskat biçimlendiricinin temel yapısı

Kaskat formant sentezi, genelde ünlü harflerin üretimi için kullanılır ve daha az kontrol parametresi kullandığı için, gerçekleşmesi paralel formant sentezine göre daha kolaydır. Fakat bu modelle, istenilenin dışında, sürtüşmeli ve patlamalı seslerin üretimi en büyük sorundur.

Paralel formant yönteminde ses aksettiriciler (resonators) Şekil 2.4’de görüldüğü üzere paralel olarak bağlanmışlardır. Bazen genizden gelen sesler için ayrıca

rezonatör de kullanılabilmektedir, ancak kademeli biçimlendiricide olduğu gibi paralel biçimlendiricide de bazı seslilerin modellenmesi problemi oluşmuştur[20].



Şekil 2.4 : Paralel biçimlendiricinin temel yapısı

Kaskat yaklaşım ile paralel yaklaşım karşılaştırıldığında, paralel sentezde her bir formant için genlik hesabı yapılması ve formant başına fazladan bir genlik hesabı yapılması gerekir. Burundan çıkan seslerin üretimi esnasında ses, burun boşluğundan geçer ve bundan dolayı burundan çıkan seslerin akustik yolu, ünlü harflere nazaran daha uzundur. Bu durumun sonucu olarak konuşulan sözcüğün bant genişliğinde ve rezonans sayısında artış olur. Bu nedenle burundan çıkan seslerin sentezi için fazladan bir rezonatör gereklidir[5].

2.6.3 Eklemeli (concatenative) sentezleme yöntemi

Anlaşılır ve doğal ses üretmenin en basit yolu, önceden sesleri kaydedip, belli formüllerle bir araya getirmektir. Ancak tüm sesleri önceden kaydetmenin belli başlı zorlukları da beraberinde gelmektedir. Örneğin üretilen ses sınırlı sayıda konuşmacıya ait olmakta ve farklı seslerin de üretilmesi isteniyorsa tüm seslerin tekrar tekrar kaydedilmesi gerekmektedir. Ayrıca, ses veritabanına yeni bir ses kaydetme işlemi başta kayıt yapılan konuşmacıya bağımlı hale gelmektedir. Bunun dışında tüm seslerin kaydedilmesi diğer yöntemlere göre çok daha fazla bellek gerektirmektedir, ama günümüzde bellek maliyetleri oldukça düştüğü için bu yöntem tekrar popüler hale gelmiştir.

Kayıtlı sesleri birbirine ekleyerek ses üretme yönteminin en önemli aşamalarından biri, doğru birim uzunluklarını tespit etmektir. Kayıtlı ses birimlerinin uzun seçilmesi

durumunda, daha doğal, daha az birleştirme noktası bulunan ve ses olayları üzerinde daha fazla kontrol sağlayan bir sonuca ulaşılır. Fakat bu avantajlarına karşın, ses birimlerini saklamak için gerekli bellek ihtiyacı, kısa ses birimleri seçmeye oranla oldukça fazladır, ayrıca çok daha fazla ses kaydı gerektirmektedir. Kayıtlı ses birimlerinin kısa seçilmesi durumunda ise, daha az bellek ihtiyacı doğmaktadır, fakat örnek toplama ve bunları etiketlendirme daha zor ve karmaşık hele gelmektedir. Günümüzde kullanılan eklemeli sistemlerde, yapılan uygulamanın türüne göre kelimeler, heceler, yarım heceler, ses birimleri (fonem), çift-sesler (diphones) ve hatta bazen üç ses (triphones) kullanılabilir[20].

Kelimeler, yazı yazarken ve sınırlı sayıda sözcük dağarcığı olan bazı mesajlaşma sistemlerinde en doğal ses birimleridir. Kelimelerin birleştirilmesi, göreceli olarak kolay olmaktadır ve kelime içindeki ses olayları ile de ayrıca uğraşmaya gerek yoktur. Fakat kelimeleri tek tek söylemekle, cümle içinde söylemek arasında oldukça büyük farklılıklar vardır. Kelimeleri tek tek söylediğimiz gibi cümle içinde söylemeye kalkıştığımız zaman doğallıktan oldukça uzak cümlelerle karşılaşabiliriz. Ayrıca, her dilde, çok sayıda bulunan özel isimlerin varlığı da, metin seslendirme sistemlerinde, ses birimi olarak kelime seçimini uygun kılmamaktadır. Dolayısıyla kelime eklemeli yöntem genellikle rezervasyon sistemi gibi kapalı sistemlerde (sabit ve az sayıda kelimenin kullanıldığı sistemlerde) uygulanmıştır. Bu gibi sistemlerde genellikle kelimeler önceden kaydedilen cümle içinden alınır, dolayısıyla ses kalitesi gerçeğe çok yakın olur. Ancak seslendirilecek metin kapalı bir sistem değilse veritabanı oldukça büyük olur ve bunu gerçekleştirmek oldukça uzun zaman alır.

Hece sayısı kelime sayılarına göre oldukça düşüktür. Fakat yine de birçok dil için bu sayı fazladır. Örneğin İngilizce için yaklaşık 15.000 hece olduğu kabul edilmektedir[33]. Türkçede ise durum daha olumludur. Türkçe kurallı bir yapıya sahip olduğu için hece sayısı çok daha düşük olmaktadır. Ancak, yabancı kökenli kelimelerde bulunan heceler de dikkate alındığında sayı 20.000'lere çıkabilmektedir. Yine de Türkçe hece tabanlı metinden ses geçme yöntemine daha uygun bir dildir. Çizelge 2.5'te Ertaş F. ve Eskidere Ö.'ye ait sözcük birimlerinin seçimi çalışması listelenmiştir[31].

Çizelge 2.5 : Sözcük birimlerinin seçimi

Birim	Miktar
Ses birimi	36
Çift-ses (diphone)	1500
Hece	20.000
Kelime	300.000

Genel olarak eklemeli yöntemde öngörülen problemler :

- Ekleme yerlerinde oluşan gürültü
- Kayıt işlemlerinin yapılabilirdiği kadar tekdüze yapılması gerekir, aksi takdirde üretilen seste bozulmalar ve çatlamalar oluşur.
- Bellek ihtiyacı (özellikle hece ve kelime yöntemlerinde seçimlerinde)
- Ses dosyalarının oluşturulması ve etiketlenmesi

Seslerin farklı zamanlarda kaydedilmesi ve aynı tekdüzelikte kaydedilmemesinden dolayı, ekleme yerlerinde ses bozukluğu ve dolayısıyla doğallıktan uzak sonuçlar oluşabilmektedir. Bu yüzden çeşitli yöntem ve algoritmalar geliştirilmiştir. En bilinen yöntem, PSOLA (Pitch Synchronous OverLap Add), France Telekom (CNET) tarafından geliştirilmiştir. Tam olarak sentezleme yöntemi olmamakla birlikte, ekleme yerlerinde daha çatlamasız geçiş sağlar, ayrıca süre ve pitch değerleri için daha iyi kontrol sağlar. Bölüm 1.3.7’de de anlatıldığı üzere TD-PSOLA, FD-PSOLA ve LP-PSOLA gibi versiyonları bulunmaktadır. PSOLA yönteminde pitch değeri seslilerde olumlu sonuçlar verirken, sessizlere uygulandığında seste gürültü meydana gelebilmektedir[34].

3. TÜRKÇE İÇİN METİN SESLENDİRME SİSTEMİ VE GERÇEKLENMESİ

3.1 Amaç

Bu çalışmadaki öncelikli amaç, Türkçe için, mümkün olduğunca doğal ve anlaşılır, metinden ses üretme sisteminin gerçekleştirilmesidir. Bunun için literatürdeki çalışmalardan ikisi göz önünde bulundurulmuş, geliştirmesi yapılmıştır. Bu çalışmalara ulama, hece geçişleri v.b. gibi dilbilimsel çalışmalar eklenmiştir. Ayrıca ses uzunluğu ve genlik değişimleriyle anlaşılabilirliğin artırılması ve istatistiksel olarak gösterilmesi hedeflenmiştir.

3.2 Tasarım

Türkçe metin seslendirme sisteminin tasarımı ve karşılaşılan problemler, diğer dillerden çok farklı değildir. Sistem tasarımını, ana hatlarıyla iki temel sınıfta kümelendirebiliriz: Metin çözümlemesi aşaması ve ses sentezi aşaması. Öncelikle her iki aşamada da karşılaşılan problemler incelenmelidir.

3.2.1 Metin çözümlemesi aşamasında karşılaşılan problemler

Metinden sese dönüştürme işleminde ilk karşılaşılan sorun girilen metnin söyleyişteki karşılığının elde edilmesidir. Bu aşama dile çok bağımlıdır ve dile özgü çözümler içermelidir.

Dönüştürme üç aşama olarak özetlenebilir; metin önışleme, doğru söyleyiş için dilbilimsel veri çıkarımı, doğru tonlama, vurgu ve süre için ölçü analizi yapılması.

3.2.1.1 Metin önışleme

Metin önışleme aşaması dile bağımlı çözümler gerektiren bir aşamadır. Öncelikle girişi yapılan metnin sözcüklerle ifade edilebilir duruma getirilmesi gerekmektedir. Yani rakamlar, sayılar, kesirler, tarihler, sıra belirten ifadeler, kısaltmalar, özel karakterler gibi yazı dilinde anlamı olan ifadeler, okunurken sarf edilen sözcüklere dönüştürülmelidir.

Örneğin, 1876 sayısı “binsekizyüzyetmişaltı” şeklinde sentezleyiciye verilmelidir. Türkçe metin seslendirme sistemi, girilen sayıyı metne çevirme konusunda kentilyon mertebesine kadar sayıları destekleyebilmektedir.

Buna 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, “2.” gibi sıra belirten ‘.’ karakterinde cümle sonu, kısaltma ya da tarih gibi bir ifade olmadığının anlaşılması ve önündeki sayıya uygun olarak “ikinci” şeklinde sentezlenmelidir.

Kısaltmalar okunuş bakımından belli kurallara sahip olmadığından karışıklık olma ihtimali yüksektir. Bu yüzden harf harf söylenmesi daha uygun olacaktır. Örneğin N.A.T.O. kısaltması birçok kişi tarafından kelime şeklinde seslendirilir, ancak aykırı sözlük kullanımı ile istenen kısaltmanın istendiği şekliyle okunması sağlanabilir.

Saat, para tutarı, ölçü, istatistik verilere ilişkin sayılar ile büyük sayılarda dönüştürme yapmak gerekir. “saat 17.30'da” ifadesi “saat onyediyüz otuzda” şeklinde, “1.500.000 lira” ifadesi “birbuçuk milyon” şeklinde okunabilir[24].

Görüldüğü gibi metin önışlemede her olası durumu doğru bir şekilde çevirmek için karmaşık kurallar tanımlamak gerekebilmektedir.

3.2.1.2 Dilbilimsel analiz

Türkçe sesçil bir dil olmasına karşın bazı istisna durumlar söz konusu olmaktadır. Örneğin aşağıdaki cümlelerde yer alan hala kelimesi,

-Annem hala gelmedi.(henüz anlamında ve yumuşak okunur)

-Babanın kızkardeşine hala denir.(sert okunur)

-Ayşe hala gelmedi.(her iki anlamda da kullanılmış olabilir)

cümlelerinde olduğu gibi okuma şekline önceden karar verilmelidir. “Annem hala gelmedi” cümlesinde “hala” sözcüğü henüz anlamında kullanılmıştır ve yumuşak okunur. “Babanın kızkardeşine hala denir” cümlesinde ise “hala” sözcüğü sert okunur. Ancak “Ayşe hala gelmedi” cümlesindeki “hala” sözcüğü her iki anlamda da kullanılmış olabilir.

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. Ancak bazı durumlar vardır ki doğal dil işleme çalışmalarında dahi doğru sonuç alınamayabilir. Sesçil olmayan dillerde bu tip problemler oldukça fazladır.

3.2.1.3 Ölçü analizi

Metnin doğru vurgu, tonlama ve uzunlukta okunabilmesi için gerekli bir adımdır. Çalışma kapsamında bu kısım üzerinde analiz sadece belli kurallar ölçüsünde kelime ve/veya cümlelerin pozisyonuna göre tanımlanmıştır. Örneğin her cümlelerin sonundaki kelime diğerlerine göre belli katsayıda daha yüksek ve uzun okunacak şekilde analiz yapılmaktadır.

3.2.2 Ses sentezi aşamasında karşılaşılan problemler

Metin analizinde olduğu gibi, sentezleme aşamasında da seçilen yönteme göre değişiklik gösteren problemler bulunmaktadır.

Mafsallı (articulatory) sentezlemede verilerin uygun şekilde kurallara dönüştürülmesi ve buna göre ses sisteminin modellenmesi çok karmaşık bir yapı gerektirmektedir. Bu karmaşık yapı gerekli olan hesap süresi ve işlemci gücünü büyük oranlarda arttıracaktır.

Biçimlendiricili (formant) sentezlemede ise biçimlendirici frekanslarını, genlik bilgilerini ve uyarım kaynağının özelliklerini kontrol edebilmemizi sağlayan kurallar oldukça fazladır. Ayrıca bu yöntemde doğallık problemleri görülmektedir. Özellikle burun-geniz kaynaklı seslerde büyük doğallık problemleri görülmektedir.

Eklemeli (concatenative) yöntemde ise gerekli olan veritabanını oluşturmak, bu veritabanında bulunan ses dosyalarını etiketlemek oldukça uzun zamanlar almaktadır. Bu veritabanının büyüklüğü bellek problemine yol açabilirken, ekleme noktalarında oluşan gürültü, anlaşılabilirliği olumsuz yönde etkileyebilmektedir.

3.3 Türkçe Metin Seslendirme Sistemi

Kullanılan sentezleme yöntemine göre metin ve dil analizi yapan modülün çıkışı da değişmektedir. Şekil 1.1’de de görüleceği üzere bu çalışmada ilk modülün girişi düz

metin iken, çıkışı fonetik seviyede bir XML katarıdır. Sentezleme aşamasında ise (ölçü, vurgu ve konuşma sentezi) girdi olarak XML katarı (metin analiz modülü de bu standartta XML katarı üretebilmektedir) alınmaktadır. Çalışmada özellikle XML format seçilmiştir ve her iki bileşenin birbirine bağımlılığı mümkün olduğunca azaltılmıştır. Her iki modül de birbiri olmadan çalışabilmektedir. Örneğin, daha kapsamlı doğal dil işleme bileşenin eklenmesiyle üretilecek olan XML katarı veya dosyası, doğrudan sentezleme bileşenine gönderilip ses dosyası üretilebilecektir.

Sentezleme bileşeninde ses veritabanının hazırlanması ve eklemeli sentezlemeye uygun şekilde birleştirme işleminin gerçekleştirilmesi yapılmaktadır. Bu kısımda üretilen ses dosyaları ve sentezleme işlemi üzerinde durulacaktır.

3.3.1 Ses dosyaları

Sesleri birçok değişik formatlı dosyalarda saklayabiliriz ancak bu projede en çok bilinen formatlardan biri olan “wav” formatı kullanılmaktadır[35]. Wav dosyasında örnekler ham veri şeklinde sıkıştırılmadan veya değiştirilmeden tutulur. 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 3.1’de gösterilmiştir.

Çizelge 3.1 : RIFF veri bölgesi (chunk) – 12 sekizli

sekizli sırası	Açıklama
0 - 3	RIFF (ASCII karakterleri şeklinde) Little Endian Şekilde paketin geri kalanının boyutu
4 - 7	
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 3.2’de gösterilmiştir.

Üçüncü veri bölgesi ise DATA’dır ve bu alanda gerçek örnekleme verileri tutulur. DATA veri bölgesi alanları Çizelge 3.3’de gösterilmiştir.

Wav dosya formatında bir örneğin aldığı değer, sesin o noktadaki enerji seviyesini vermektedir. Bu değer mutlak değeri sesin şiddetiyle yakından ilgilidir, yani örneğin mutlak değerindeki artış, sesin gücündeki artışı vermektedir.

Çizelge 3.2 : FORMAT veri bölgesi (chunk) – 24 sekizli

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
	Örnekteki sekizli anlamı :
	1 = 8 bit mono,
	2 = 8 bit stereo veya 16 bit mono,
20 - 21	4 = 16 bit stereo
22 - 23	Örnekteki bit sayısı

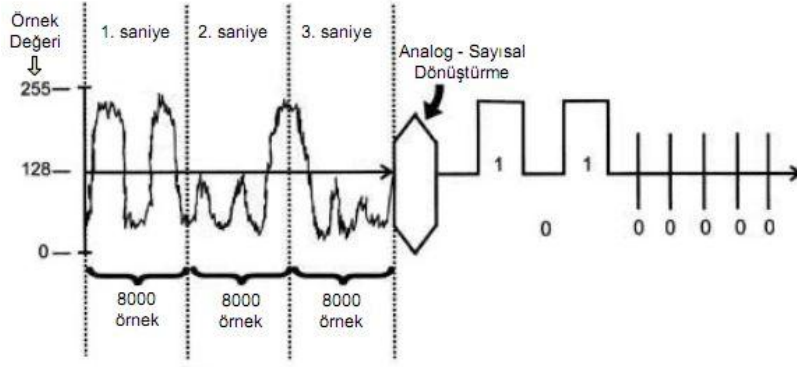
Çizelge 3.3 : DATA veri bölgesi (chunk)

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

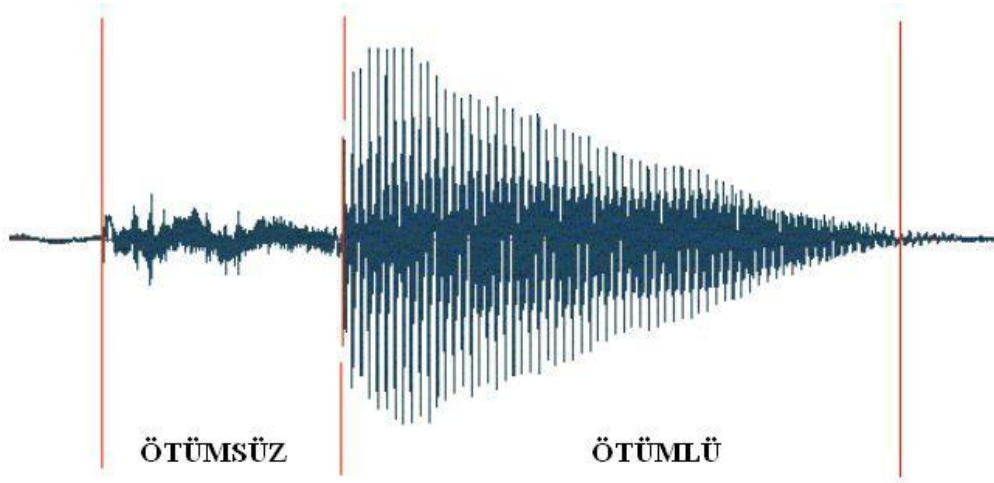
Ses, analog bir işarettir ve kodlanabilmesi için sayısal işarete dönüştürülmesi gerekmektedir. Bu dönüştürme örnekleme ile yapılır. Sinyalden bir saniyede birçok kez örnek alınarak dalganın yüksekliği ölçülür. Bu tip modülasyona “Darbe Kod Modülasyonu” adı verilir. 8 bitlik örneklemede sesin analog seviyeleri 255 adet sayısal seviyeye dönüştürülür. Örneğin saniyede 8000 örnek alınırsa bir saniyelik işaret $8000 \times 8 = 64000$ bit ile ifade edilmiş olacaktır. Aynı şekilde 16 bitlik örneklemede 65536 adet sayısal seviye bulunmaktadır. Örneğin, 16 bitlik örneklemede ilk bit işaret biti olup örnek, -32768 – 32767 arasında değer alabilmektedir. Çalışmanın gerçekleştirildiği Java programlama dilinde “int” veri tipi 64 bit olduğu için işaret biti en anlamlı bit olan 64. bittir. Bu yüzden yapılan çalışmada ses ile ilgili değişikliklerde işaret bitinin taşınması da gerekmiştir.

Şekil 3.1’de ses sinyalinin örneklenmesi özetlenmiştir[5].

İnsan konuşmasında temelde iki farklı ses vardır. Bunlar ötümlü ve ötümsüz olarak adlandırılabilir. Ötümlü (*voiced*) ses zaman çerçevesinde incelenirse periyodik karakteristik gösterdiği görülebilir. Diğer taraftan ötümsüz (*unvoiced*) ses periyodik olmayan bir davranış sergilemektedir. Şekil 3.2’de ötümlü ve ötümsüz seslerin dalga şekli görülmektedir[5].

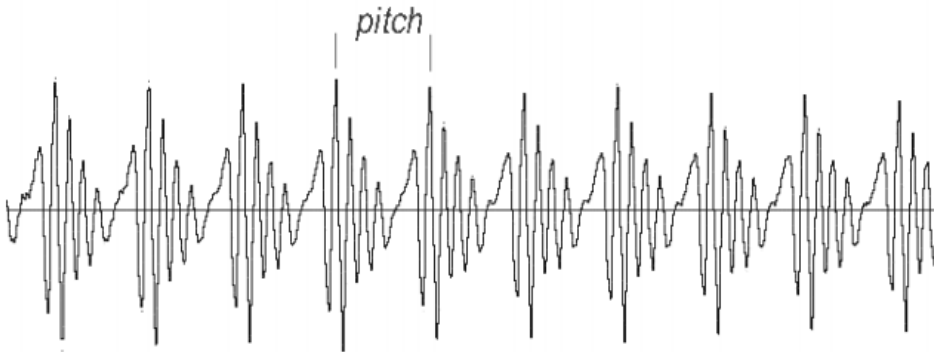


Şekil 3.1 : Ses sinyalinin örneklenmesi



Şekil 3.2 : Ötümlü ve ötümsüz ses

“Pitch”, ses verisinin bir periyoduna verilen isimdir ve sadece periyodik olan ötümlü kısımda söz konusu olabilmektedir. Ötümsüz kısım periyodik olmadığı için “Pitch” değerlerinden söz edilemez. “Pitch” değeri belirli bir ses parçasındaki örnek sayısının periyot sayısına bölümü ile hesaplanabilir. Örneğin, bir ses dosyasında 1800 adet örnek ve 10 periyot varsa “pitch” değeri $1800 / 10 = 180$ olarak hesaplanır. Periyodik ses sinyali ve “pitch” periyodu Şekil 3.3’de gösterilmiştir[5].



Şekil 3.3 : Periyodik ses sinyali ve “pitch” periyodu

3.3.2 Ses dosyalarının (veritabanı) hazırlanması

Ses veritabanı hazırlama aşaması, eklemeli sentezleme yöntemine göre metinden ses üreten sistemlerde en önemli adımdır. Sistemin çalışabilmesi için gerekli olan ses birimlerinin, önceden kayıt edilmiş olması gerekmektedir. Ancak tüm seslerin tek tek kaydedilerek etiketlenmesi oldukça fazla zaman ve çaba gerektirmektedir. Bu yüzden seslerin dış kaynaklardan elde edilmesi yönünde çalışmalar yapılmıştır.

Öncelikli olarak TDK sesli sözlük veritabanı incelenerek, çift-ses ekleme (diphone concatenation) ve hece ekleme (syllable concatenation) yöntemleri test edilmiştir. TDK sesli sözlük veritabanındaki sesler arasında erkek ve kadın sesleri karışık olarak yer almaktadır. Ayrıca sesler farklı vurgu ve tonlamayla seslendirilmiştir. Eklemeli yöntemlerde seslerin mümkün olduğunca monoton ve ritimsiz (vurgu ve tonlamadan uzak) olması gerekmektedir. Bu problemlerden ötürü TDK sesli sözlükten vazgeçilmiştir.

Festival ses sentezleme sistemi veya Mbrola kullanıldığında seçilen ses birimi çift-ses olduğu için buna uygun şekilde bütün çift-seslerin oluşturulmuş olması gerekmektedir. Hece tabanlı bir sistemde ise tüm hece listesinin tamamının kayıt edilmesi gerekmektedir. Bu çalışmada Türkçe için çift-ses ve hece eklemeli sistemler gerçekleştirilmiştir. Dolayısıyla Türkçede bulunan tüm çift-ses ve hecelerin önceden kayıt edilmesi problemi mevcuttur. Çalışmanın hedeflerinden birisi de, ses kaydının titiz bir çalışma sonucunda, düzgün ve monoton kaydedilmesi ile, sonucun da anlaşılır ve doğal olabileceğinin gösterilmesidir.

Her bir çift-ses ve hecenin önceden kaydedilmesi ve etiketlenmesi çok uzun bir süreçtir. Bu yüzden; öncelikle, küçük bir veritabanı 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 sağlanmıştır.

3.3.2.1 Türkçe sesçil (fonetik) alfabesi

Çift-ses eklemeli yöntem ile ses üretilen Festival ve Mbrola sistemleri incelenmiş, Mbrola ile tüm seslerin otomatik olarak üretilmesi hedeflenmiştir. Mbrola ile ses üretimi yapılabilmesi için fonetik işaret gerekmektedir. Mbrola'nın desteklediği ve desteklemediği sesleri içeren Türkçe sesçil alfabesi Çizelge 3.4'de

gösterilmiştir. Bu alfabe oluşturulurken çizelgede SAMPA ve MBROLA'nın kullanmış olduğu fonetik işaretler de dikkate alınmıştır[36]. Bazı sesler MBROLA veritabanında yer almamaktadır (Çizelge 3.4'de MBROLA sütununda “-“ olarak belirtilmiştir).

Bunun dışında Çizelge 3.4'te, her bir ses için; ses dosyalarında bulmak üzere “dosya adı karakteri” alanına ihtiyaç duyulmuştur. Bu alan her bir hece için hangi ses dosyasının eklenmesi gerektiğini bulmak için eklenmiştir.

Türkçe'de “ğ” sesi genellikle, peşine geldiği seslinin daha uzun okunması için kullanılır. Örneğin, “eğer” sözcüğünde bulunan “ğ” sesi, kendinden önce gelen “e” sesinin biraz daha uzun okunmasını sağlar.

Çizelge 3.4'de bulunmayan bir ses de “renk” sözcüğünde bulunan “e” sesidir. “keçi” veya “yemek” sözcüklerindeki “e” sesine göre daha yayvan söylenmektedir.

Çizelge 3.4 : Türkçe sesçil alfabesi

Türkçe Sesçil Alfabe				
Harf	Örnek kelime	SAMPA	MBROLA	Dosya adı karakteri
a	kal, aşk	a	a	a
b	balık, bataklık	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, gavur	gj	g	g
g	karga, gaga	g	-	-
ğ	sağ (sol tersi), sağır	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

Çizelge 3.4 (devam) : Türkçe sesçil alfabesi

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 kullanılan heceler, tüm çift-sesleri (diphone) kapsamaktadır. Dolayısıyla, Türkçede kullanılan tüm hecelere ait ses dosyalarının üretilmesi çalışma için yeterli olacaktır. Çizelge 3.4 dikkate alındığı takdirde, Mbrola kullanılarak üretilcek ses dosyalarında 8 adet sesli, 22 adet sessiz bulunmaktadır.

3.3.2.2 Türkçede kullanılan hece tipleri

Türkçe sesçil alfabesi dikkate alındığında Mbrola ile üretilmesi gereken sekiz farklı hece türü bulunmaktadır. Bunlar en az bir, en fazla dört harften oluşur. Aslında öz Türkçede altı farklı hece tipi bulunmaktadır, ancak günümüzde diğer dillerden gelen ve dilimizce benimsenmiş birçok kelime bulunmaktadır. Diğer hece tipleri de desteklenmediği takdirde çoğu metin seslendirmesinde problem yaşanacaktır. Bu yüzden çalışma kapsamına yabancı kökenli sözcükler de 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 (Örneğin al, an, et, üç, öl). Matematiksel olarak $21 \times 8 = 168$ adet hece bulunmaktadır. Ancak Çizelge

3.4’de Mbrola’nın desteklediği 22 adet sessiz olduğu için $22 \times 8 = 176$ adet ses dosyası oluşturulabilmektedir.

- **VC** tipinde heceler

Sesli harf + sessiz harf şeklinde oluşan hecelerdir (Örneğin ba, ce, zi, gü). CV hece tipinde olduğu gibi $8 \times 22 = 176$ adet ses dosyası oluşturulabilmektedir.

- **VCC** tipinde heceler

Sesli harf + sessiz harf + sessiz harf şeklinde oluşan hecelerdir (ilk, ürk, ast). Matematiksel olarak $8 \times 22 \times 22 = 3872$ 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”. Yabancı kökenli kelimelerde de bulunabilen “rz” sessizleri de eklendiği takdirde Çizelge 3.5 ortaya çıkmaktadır.

Çizelge 3.5 : 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	Rüş
yt	Hayt

Daha çok sesleniş biçiminde sözcüklerde kullanılan (ör. “hey””, “höy””) ve “yt” ile biten heceler de eklenmelidir. Ayrıca “-l” sessizi, “lale” ve “halı” sözcüklerinde farklı seslendirildiği için bu kuralda “-l” ile ilgili sesler de çoklanmalıdır. ($4 + 18 = 22$)

$8 \times 22 = 176$ olmak üzere toplam 152 ses dosyası bulunmaktadır.

- **CVC** tipinde heceler

Sessiz harf + sesli harf + sessiz harf şeklinde oluşan hecelerdir (kal, tek, bit). $22 \times 8 \times 22 = 3872$ 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 $22 \times 8 \times 22 = 3872$ adet ses dosyası bulunmaktadır.

- **CCV** tipinde heceler

Sessiz harf + sessiz harf + sesli harf şeklinde oluşan hecelerdir (tra, spo, gri). Yabancı kökenli kelimelerde bulunabilen CCV tipi hece yapısı sadece sözcük başlarında bulunabilir. Matematiksel olarak $22 \times 22 \times 8 = 3872$ adet ses dosyası oluşturulmalıdır ancak yabancı kökenli sözcüklerin başında bulunabilen bu hece tipi sadece Çizelge 3.6’da görüldüğü üzere (br, bl, dr, fr, gl, gr, hr, kl, kr, pl, ps, tr) listesindeki iki sessizlerle başlayabilirler. Dolayısıyla CCV hece tipinde $12 \times 8 = 96$ adet ses dosyası gerekmektedir.

Çizelge 3.6 : 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, trolleybüs

- **CCVC** tipinde heceler

CCV hece tipine benzer bir şekilde yabancı kökenli kelimelerde 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 hecelerın sayısı sadece 96 iken CCVC tipindeki heceler $12 \times 8 \times 22 = 2112$ adet ses dosyası gerektirmektedir. Dolayısıyla CCVC tipindeki hecelerde bu indirgeme aktif hale getirilmiş, CCV tipindeki hecelerde ses üretme yoluna gidilmiştir.

Hece eklemeli yöntemde, Türkçe metin seslendirme sistemi için yaratılması gereken ses dosyaları toplamı (en çok) Çizelge 3.7’te gösterilmiştir.

Çizelge 3.7 : 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	$21 \times 8 = 168$	$22 \times 8 = 176$
3	VC	ba, ce, zi, gü	$8 \times 21 = 168$	$8 \times 22 = 176$
4	VCC	ilk,ürk, ast	$8 \times 22 \times 22 = 3872$	$8 \times 22 = 176$
5	CVC	kal, tek, bit	$22 \times 8 \times 22 = 3872$	$22 \times 8 \times 22 = 3872$
6	CVCC	türk, sark, dört	$22 \times 8 \times 22 \times 22 = 85184$	$22 \times 8 \times 22 = 3872$
7	CCV	tra, spo, gri	$22 \times 22 \times 8 = 3872$	$12 \times 8 = 96$
8	CCVC	tren, kral	$22 \times 22 \times 8 \times 22 = 85184$	Hece indirgeme ile 0
	Toplam		182328	8376

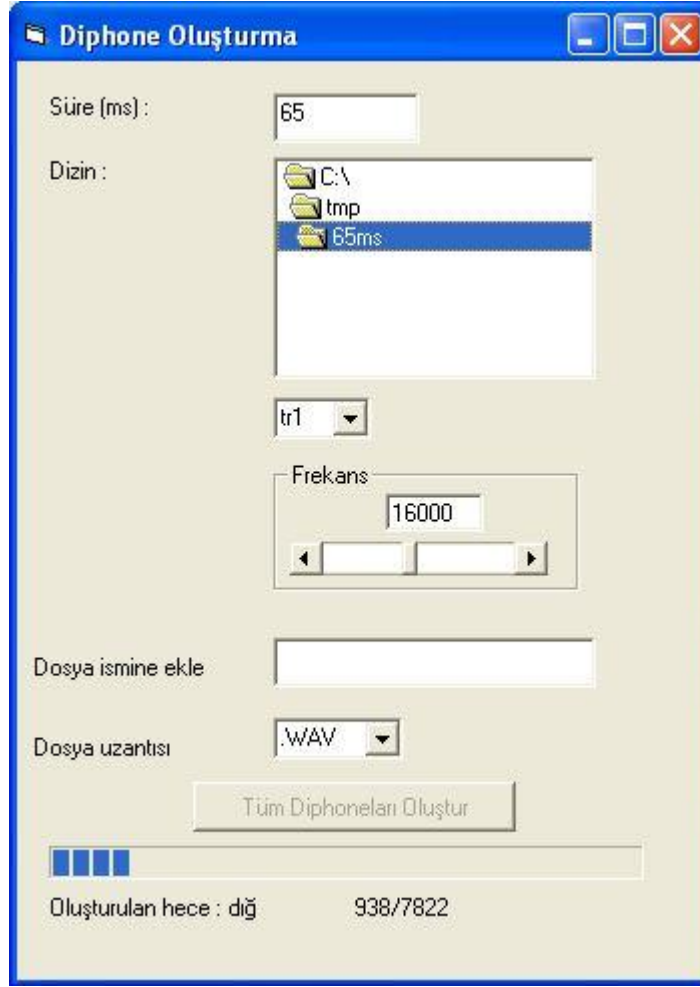
3.3.2.3 Mbrola kullanılarak ses veritabanının otomatik olarak oluşturulması

Belirlenen kurallara göre ses veritabanı Mbrola aracı kullanılarak oluşturulmuştur. Bunun için Mbrola aracının kullanmış olduğu dinamik kütüphaneler (dll) incelenmiş ve nasıl kullanılabileceği araştırması yapılmıştır.

Mbrola uygulamasının Windows sürümü Microsoft standardı olan bu kütüphaneleri (dll) kullandığı için ve asıl kullanılacak olan Türkçe Metin Seslendirme sistemi uygulaması Java ile programlandığı için, bu kütüphanelerin Java üzerinden de çağrılabilmesi gerekmektedir. Bu yüzden hazırlanan MbrDll isimli Visual C++ projesi ile Java'nın kullanabileceği şekilde yeni dll kütüphanelerinin oluşturulması sağlanmıştır[39]. Yeni oluşturulan kütüphaneler sonucunda tüm sesler otomatik olarak Şekil 3.4'te görüldüğü gibi program vasıtasıyla kolayca üretilebilmiştir.

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, kelimelerin oluşturulmasında da kullanılacağı için ortalama uzunlukta heceler seçilmelidir. Bu yüzden, normal konuşmada her ses için ortalama süre 65 milisaniye olarak ele alınmış ve sesler bu doğrultuda oluşturulmuştur.

Sesler tek kanal (mono), 44100 örnekleme oranı ile PCM (darbe kod modülasyonu) ve 16 bitlik örnekler olacak şekilde oluşturulmuştur.



Şekil 3.4 : Mbrola kullanarak ses veritabanının otomatik olarak oluşturulması

Girdi olarak alınan metnin değişik ses veritabanları tarafından seslendirilebilmesine yönelik çalışma da tamamlanmıştır. Bu doğrultuda, üretilen ses dosyalarını içeren ses veritabanı tanımlaması Şekil 3.5'te görülen ekran aracılığıyla tanımlanabilmektedir. Ayrıca bu ekran üzerinden güncelleme ve silme işlemleri de yapılabilmektedir.

3.3.2.4 Oluşturulan seslerin genliklerinin dengelenmesi

Otomatik olarak oluşturulan dosyalar arasında dengelenmemiş -farklı genlikte- ses dosyaları mevcut olabileceğinden ekleme yerlerinde seste çatlama oluşmaması amacıyla, tüm dosyalar üzerinde bir algoritma çalıştırılarak seslerin dengelenmesi hedeflenmiştir.

Ses veritabanı işlemleri

Ses veritabanı

Ses veritabanı adı: tr1_65

Ses ana dizini: C:/tmp/65ms/

Acıklama: Türkçe Erkek (65 ms)

Frekans: 16000

Kod: tr1

Ekle **Sil**

Veritabanı adı	Ses ana dizini	Veritabanı açıklaması	Frekans	Kod
tr1	C:/tmp/male/	Türkçe Erkek	16000	tr1
tr2	C:/tmp/femal...	Türkçe Kadın	23000	tr2
tr1_65	C:/tmp/65ms/	Türkçe Erkek (65 ms)	16000	tr1

Şekil 3.5 : Oluşturulan ses dosyalarını içeren veritabanının uygulamaya tanıtılması

Bu dengeleme çalışması basitçe aşağıdaki adımlardan oluşmaktadır :

- Tüm ses dosyaları taranarak, en yüksek, en düşük ve ortalama genlik değerleri belirlenmiştir.
- Tüm ses dosyaları taranarak, bir üst adımda belirlenmiş olan en yüksek, en düşük ve ortalama genlik değerlerine göre karşılaştırılarak genlik değerleri dengelenmiştir. Bunun için üç farklı yöntem kullanılıp denenmiştir.
 - 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.
 - 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.

- 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.

3.3.3 Türkçe Metin Seslendirme Sisteminin Gerçeklenmesi

Genel olarak Türkçe metin seslendirme sistemini iki ana bileşende inceleyebiliriz. Her bileşen kolayca ayrılabilir (gevşek bağlaşımlı) ve dış sistemlerle tümleştirilebilir şekilde tasarlanmıştır.

Şekil 1.1’den de görüleceği üzere, ilk bileşen (metin işleme bileşeni), girilen metnin belli başlı kurallardan geçirilerek sistemin anlayacağı dile (bu çalışmada XML olarak belirlenmiştir) çevrilmesini sağlar. Diğer bileşen ise (ses işleme bileşeni), girdi olarak fonetik seviyede XML alıp, ses dosyası üretir. Bu bileşen, ilk bileşen olmadan, farklı sistemlerden XML katarı veya dosyası ile beslenerek ses üretilebilir şekilde tasarlanmıştır. Aynı şekilde ilk bileşen de farklı sistemlerden metin katarı alıp, XML katarı üretebilir şekilde tasarlanmıştır.

Her iki bileşeni de kullanarak kullanıcıya arayüz hazırlanmıştır ve bu arayüz aracılığıyla kullanıcının girmiş olduğu metin ses dosyasına dönüştürülüp, kullanıcıya dinletilmektedir. Şekil 3.6’da bu sistemin temel akışı görülmektedir.

3.3.3.1 Metin işleme bileşeni

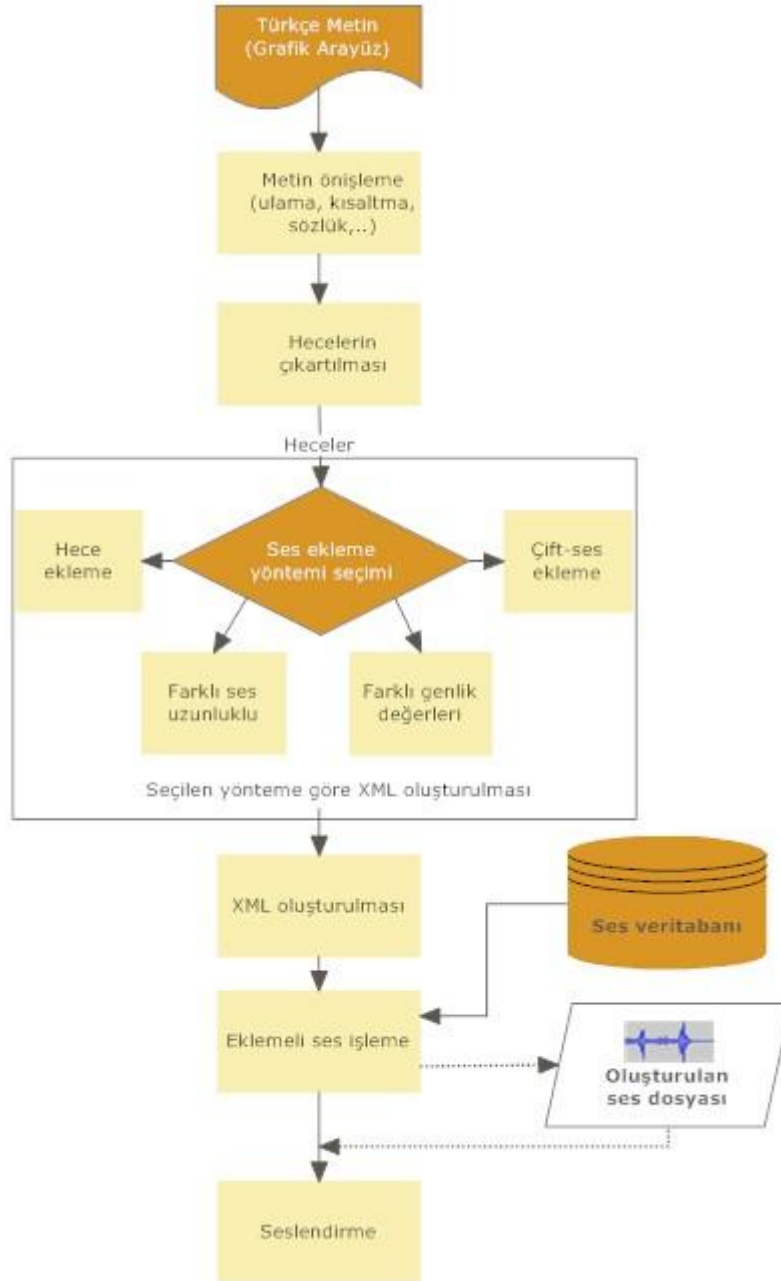
Türkçe Metin Seslendirme sisteminin ilk bileşenidir. Bu bileşende, girilen metin ikinci bileşenin (ses işleme bileşeni) istediği XML biçiminde katar üretir. Temel olarak bu bileşende yapılan işlemler aşağıdaki adımlarla özetlenebilir.

1. Girdi olarak gelen metin, içinde bulunan noktalama işaretlerine ve boşluklara göre ağaç yapısına dönüştürülür. Bu ağaç yapısında metin, cümle, kelime, hece, ses ve noktalama işaretleri bulunmaktadır.
2. Her bir cümle içinde bulunan kelimeler 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.

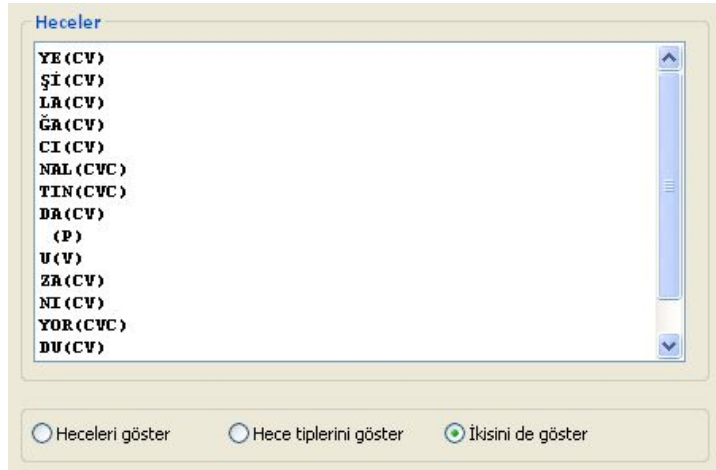


Şekil 3.6 : Türkçe Metin Seslendirme sisteminin temel akışı

(C + E [1]) yapısına, yanında bir ünlü olmayan ünsüze rastlandığında veya art arda gelen ünsüzlerden anlam çıkartılamadığında gelir. 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 karşılaşılmıştır. 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 kelimeler hecelere ayrılma aşamasından sonra “ne ta-ne tam-sa-yı” ve “Pe Te Te ku-ru-mu” şeklini alır.

4. Şekil 3.8’de görüleceği üzere hece parçalama algoritma sonrası heceler listelenir. Bu listenin; heceler, hece tipleri veya her ikisi birden olacak şekilde üç farklı gösterimi mevcuttur. (C : sessiz, V : sesli, P : noktalama ve duraklama işaretleri)



Şekil 3.8 : Hecelere parçalama sonrasında hecelerin gösterimi

5. Örneği Şekil 3.9’da görülen yapıda XML katarı oluşturulur. Burada görülen vurgu katsayısı ve süre bilgileri “Çift-ses eklemeli yöntem” ve “Hece eklemeli yöntem” bölümlerinde anlatılacaktır.

```

<?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">
        <ses sampa="G" sure="60"/>
        <ses sampa="a" sure="90"/>
      </hece>
      <hece vurguKatsayisi="1">
        <ses sampa="dZ" sure="60"/>
        <ses sampa="@" sure="90"/>
      </hece>
      <hece vurguKatsayisi="1">
        <ses sampa="n" sure="60"/>
        <ses sampa="a" sure="90"/>
        <ses sampa="l" sure="60"/>
      </hece>
      <hece vurguKatsayisi="1">
        <ses sampa="t" sure="60"/>
        <ses sampa="@" sure="90"/>
        <ses sampa="n" sure="60"/>
      </hece>
      <hece vurguKatsayisi="1.4">
        <ses sampa="d" sure="90"/>
        <ses sampa="a" sure="120"/>
      </hece>
    </kelime>
    <bekle sure="50"/>
    <kelime vurguKatsayisi="1">
      <hece vurguKatsayisi="1">
        <ses sampa="u" sure="90"/>
      </hece>
      <hece vurguKatsayisi="1">
        <ses sampa="z" sure="60"/>
        <ses sampa="a" sure="90"/>
      </hece>
      <hece vurguKatsayisi="1">
        <ses sampa="n" sure="60"/>
        <ses sampa="@" sure="90"/>
      </hece>
      <hece vurguKatsayisi="1">
        <ses sampa="j" sure="60"/>
        <ses sampa="o" sure="90"/>
        <ses sampa="r" sure="60"/>
      </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>

```

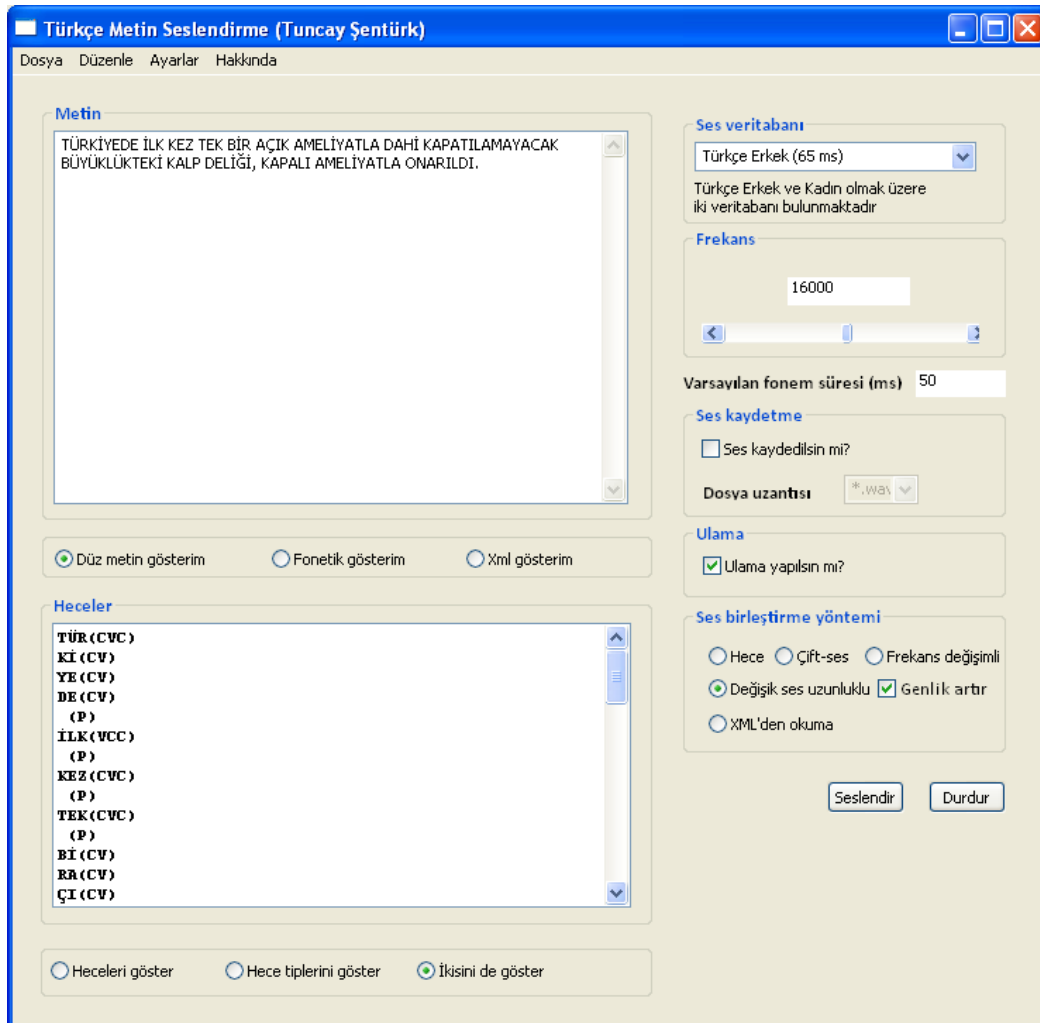
Şekil 3.9 : Metin işleme birimince oluşturulan XML örneği

3.3.3.2 Ses işleme bileşeni

Türkçe Metin Sestlendirme Sistemi gerçekleştirirken birçok yöntem denenmiştir. Amaç, bu yöntemlerin her birisinin parametrik olarak denenebilmesi ve ölçülmesidir. Denenen ve istatistiksel bir şekilde başarı ölçümü yapılan bu yöntemleri aşağıdaki şekilde listeleyebiliriz :

- Çift-ses eklemeli yöntem
- Hece eklemeli yöntem
- Farklı hece uzunlukları ile eklemeli yöntem
- 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 3.10’da görüldüğü üzere, seçilerek çalıştırılabilir şekilde tasarlanmıştır.



Şekil 3.10 : Türkçe Metin Sestlendirme uygulaması ekran görüntüsü

Çift-ses (diphone) eklemeli yöntem

Anlaşılır konuşma frekans bandının 5 kHz ve örnekleme frekansının, örnekleme teoremi uyarınca, 16 kHz olduğu ortalama kayıt süresi 200 msn olan 16 bitlik yaklaşık 8000 hecenin, ortalama 1 sn'lik kaydı için yaklaşık 50MB saklama alanı gerekmektedir ($2 \times 16000 \times 0.2 \times 8000 / (1024 \times 1024)$). Yapılan otomatik ses üretiminden sonra bu değerin 55 MB olduğu gözlemlenmiştir. Günümüzde bu değer pek önemli olmasa 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 veritabanı hazırlama işleminin oldukça uzun uğraşlar gerektirmesi dolayısıyla çift-ses eklemeli yöntem de çalışmaya dahil edilmiştir.

Girilen metin içinde bulunan hecelerin çift-sesler birlikteliğine dönüştürülmesi amacıyla Çizelge 3.8 hazırlanmıştır.

Çizelge 3.8 : Hecelerin uygun çift-seslere bölünmesi

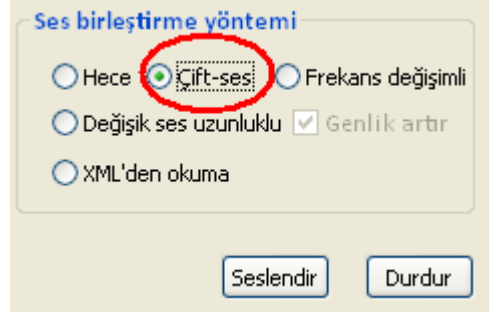
Hece tipi	Uygun çift-sesler
(V)	V + (bir sonraki hece C ile başlıyor ise bekle)
(VC)	V + VC
(CV)	CV + V + (bir sonraki hece C ile başlıyor ise bekle)
(VCC)	V + VC + C
(CVC)	CV + V + VC + C
(CVCC)	CV + V + VC + C

V-C hece geçişindeki tıkanıklık bekleme işlemi ile giderilmektedir. Bu tıkanıklık, ünlüyle biten heceden sonra ünsüzle başlayan bir hece gelmesi durumunda ortaya çıkar ve sentezde istenmeyen duraklamalara neden olur. Bunun nedeni, bu geçişteki ünlünün yeteri kadar kısa tutulamamasıdır.

Önlem olarak kısa geçişbirim kaydetmek yerine uzun geçişbirimde süre denetimi yapmak yoluna gidilmiştir. Böylece daha ünlü sonlanmadan ünsüz başlatılır ve tıkanma yerine kısa süreli ötümlülük sağlanır.

Her ünlünün aynı uzunlukta olmadığı gözönüne alınarak tüm ünlüler için farklı süre belirlenmiştir. Bu süre, deneysel olarak en az tıkanmanın olduğu durum dikkate alınarak bulunmuştur.

Şekil 3.11'de, uygulamada çift-ses eklemeli yöntemin nasıl seçileceği gösterilmiştir.



Şekil 3.11 : Uygulamada çift-ses eklemeli yöntem seçeneği

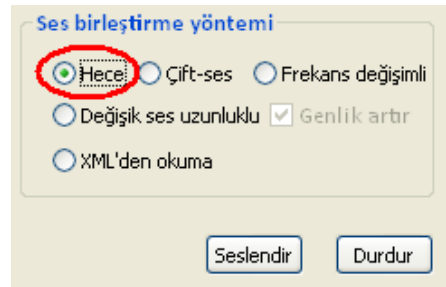
Hece eklemeli yöntem

Hece ekleme yönteminin, çift-ses ekleme yöntemine göre, daha az ekleme yeri olmasından ötürü, daha doğal sonuçlar ürettiği belirtilmektedir, ancak yakın geçmişte disk, bellek ve işlemci maliyeti problem olarak görülürken; günümüzde gelişen teknoloji sayesinde bu sıkıntı aşılmıştır.

Çizelge 3.7’den görüleceği üzere Türkçede kullanılan ve üretilmesi gereken ses dosyalarının (yabancı kökenli kelimelerden gelen heceler dahil olmak üzere) toplamı yaklaşık olarak 8000’dir. Bu ses dosyalarının sabit diskte tuttuğu alanın 55 MB civarında olduğu da gözönünde bulundurulursa Türkçe Metin Seslendirme sistemi için hece eklemeli yöntemin kullanımı daha uygun hale gelmektedir.

Türkçe Metin Seslendirme uygulaması, çift-ses eklemeli yöntemde olduğu gibi hece eklemeli yöntemi de desteklemektedir. Bunun için Şekil 3.12’de görüldüğü gibi “Ses birleştirme yöntemi” panelinde, “Hece” seçiminin yapılması gerekmektedir.

Yine, 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 veritabanları da bulunmaktadır.



Şekil 3.12 : Uygulamada hece eklemeli yöntem seçeneği

Farklı hece uzunlukları ile eklemeli yöntem

Hece eklemeli yöntemde tüm ses uzunlukları sabit olarak değerlendirilmiştir. Anlaşılır fakat doğallıktan uzak sonuçlar ortaya çıkmıştır. Doğallığı biraz daha arttırabilmek için üzerinde durulması gereken konulardan ikisi vurgu ve uzatmalardır. Farklı hece uzunlukları ile eklemeli yöntemde ses sürelerinde uzatmalar yapılarak doğallığı arttırma yoluna gidilmiştir.

Normal konuşma kaydedilerek yapılan çalışma sonucunda seslerin cümle içindeki yerlerine göre oldukça değişken uzunlukta olabildikleri tespit edilmiştir. Ancak bu çalışma kapsamında genel kural belirlenmiş ve bunun dışındaki ses uzunlukları (sesleniş, virgölün yeri, soru cümleleri v.b.) dikkate alınmamıştır. Belirlenen genel kurala göre hecelerde bulunan sessizler 60 ms, sesliler ise 90 ms olacak şekilde tanımlanmıştır. Ancak hece, kelimenin 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. Şekil 3.13'te “altında” kelimesi için üretilen XML katarı görülmektedir. Ses süreleri sarı içine alınarak belirginleştirilmiştir.

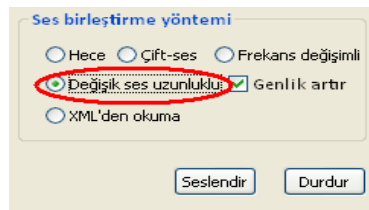
```
<kelime vurguKatsayisi="1">
  <hece vurguKatsayisi="1">
    <ses sampa="a" sure="90"/>
    <ses sampa="l" sure="60"/>
  </hece>

  <hece vurguKatsayisi="1">
    <ses sampa="t" sure="60"/>
    <ses sampa="@" sure="90"/>
    <ses sampa="n" sure="60"/>
  </hece>

  <hece vurguKatsayisi="1.4">
    <ses sampa="d" sure="90"/>
    <ses sampa="a" sure="120"/>
  </hece>
</kelime>
```

Şekil 3.13 : Sesler için süre bilgisinin XML belgesinde gösterimi

“Ses birleştirme yöntemi” panelinde bulunan “Değişik ses uzunluklu” seçimi ile bu yöntem aktif hale getirilebilmektedir (Şekil 3.14). Bu yöntem, genlik arttırma işlemi ile iç içe çalışmaktadır ve bu konuda yapılan çalışmalar daha sonra anlatılacaktır.



Şekil 3.14 : Değişik ses uzunluklu hece eklemeli yöntem seçeneği

Farklı genlik değerleri ile eklemeli yöntem

Türkçede doğallığı arttıran etkenlerden bir diğeri de vurgudur. Türkçe sözcüklerde, sözcük içinde ve cümle içinde vurgunun nerede yapılacağı yaklaşık olarak bellidir; ancak konuşma dilinde vurgunun yeri değişebilir[23]. Bu çalışmada sadece belli olan yöntem denenmeye çalışılmış ve asıl çalışmalar doğal dil işleme bileşenine bırakılmıştır.

Vurgu, genlik veya frekans değişimi ile verilebilir. Türkçe Metin Seslendirme çalışmasında her ikisi üzerinde de durulmuştur; ancak frekans değişikliği, belli kurala göre olmaksızın, konuşma sırasında verilmesi istenen vurguya dayanmaktadır. Bu yüzden yapılan uygulamada sadece genlik değişimi ile vurgu verilmeye çalışılmıştır. Frekans değişikliği ile vurgu ve doğallık verilmesi ise programatik yoldan yapılmamış, sadece alt çalışma olarak tamamlanmıştır.

Türkçede vurgu Bölüm 2.4’de ele alınmıştır ancak Türkçe Metin Seslendirme uygulamasında sadece kelimenin ve cümlenin son hecesinde vurgu olduğu yönünde geliştirme yapılmıştır. Bu kapsamda, parametrik olarak değiştirilebilen bir katsayı ayarlanarak, sözcüğün son hecesi ile cümlenin son kelimesinin bu katsayı kullanılacak şekilde, genlik değerinin düzenlenmesi çalışması yapılmıştır. Şekil 3.15’te görüleceği üzere “altında” kelimesi içinde bulunan son hece olan “da” hecesinin daha yüksek genlikle seslendirilmesi sağlanmıştır.

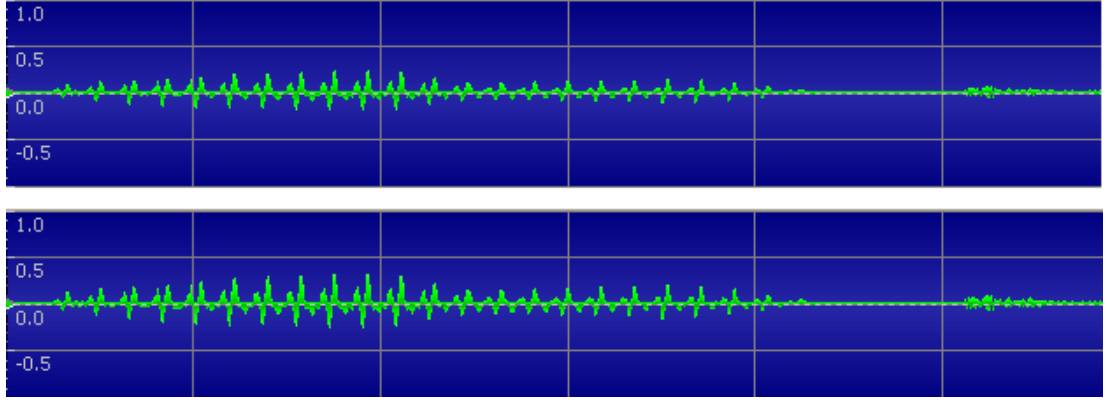
```
<kelime vurguKatsayisi="1">
  <hece vurguKatsayisi="1">
    <ses sampa="a" sure="90"/>
    <ses sampa="l" sure="60"/>
  </hece>

  <hece vurguKatsayisi="1">
    <ses sampa="t" sure="60"/>
    <ses sampa="@" sure="90"/>
    <ses sampa="n" sure="60"/>
  </hece>

  <hece vurguKatsayisi="1.4">
    <ses sampa="d" sure="90"/>
    <ses sampa="a" sure="120"/>
  </hece>
</kelime>
```

Şekil 3.15 : Hece veya kelimeler için vurgu katsayısının XML belgesinde gösterimi

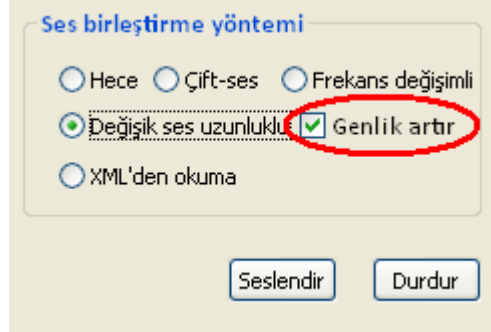
Örnek olarak Şekil 3.16’da “ilk” hecesinin normal genlikli ve 1.4 katsayısı ile genliğinin yükseltilmiş hali görüntülenmektedir.



Şekil 3.16 : “ilk” hecesinin normal ve 1.4 katsayısı ile genliğinin arttırılmış hali

16 bit örnekleme kullanıldığı için, genliğin arttırılması aşamasında 64 bit işaretli veri tipinden 16 bit işaretli veri tipine dönüşüm yapmak gerekmektedir. Bu yüzden alt seviyede programatik işlemler yapılmıştır.

Türkçe Metin Seslendirme uygulamasında, “Ses birleştirme yöntemi” panelinde bulunan “Genlik artır” seçimi ile; genlik arttırarak vurgu verme yöntemi aktif hale getirilebilmektedir (Şekil 3.17).



Şekil 3.17 : Uygulamada farklı genlik değerleri ile eklemeli yöntem seçeneği

XML'den okuma

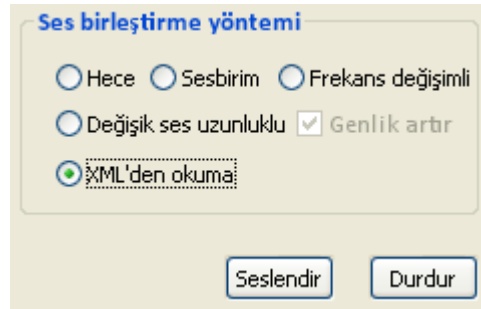
Tüm yöntemlerin yanı sıra, bir de doğrudan XML'den okuma seçeneği eklenmiştir. Bu seçenek, eklenecek olan doğal dil işleme bileşeninin üreteceği XML'i doğrudan sese dönüştürebilecek şekilde tasarlanmıştır. Şekil 3.8'de örnek XML'de gösterildiği gibi aşağıdaki özellikler bulunabilmektedir.

- Metin (<metin>) içinde bulunacak duraksamalar (ms cinsinden), noktalama işaretleri (<bekle>) ile belirtilir.

- Cümle (<cumle>) içinde bulunan tüm kelimeler (<kelime>), ve bu kelimeler arası duraksamalar (<bekle>), kelimelere ait vurgular (“vurguKatsayisi”) ile belirtilir.
- Kelimeler 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.

Türkçe Metin Seslendirme uygulamasında hazırlanmış olan XML’den ses üretme işlemi Şekil 3.18’de gösterildiği üzere, “Ses birleştirme yöntemi” panelinde bulunan “XML’den okuma” seçeneğinden yapılabilmektedir.

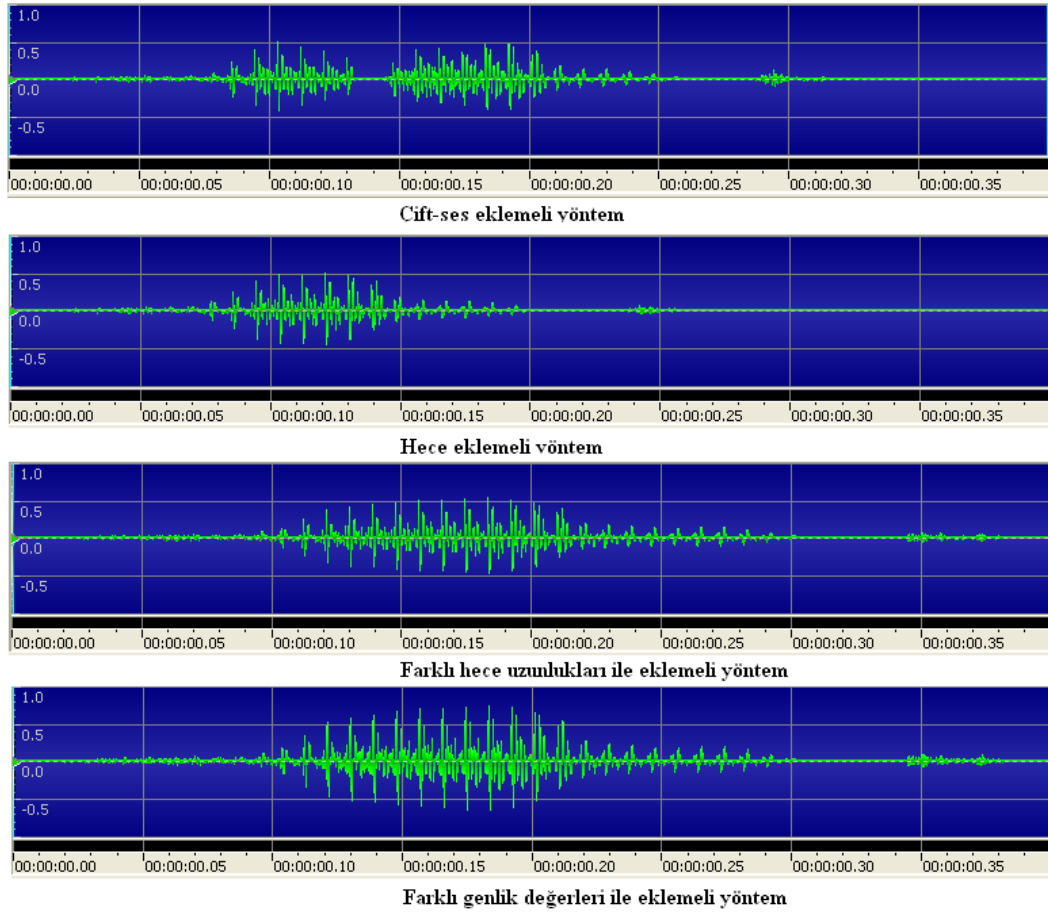
XML’den okuma seçeneği ile, seslerdeki uzatma, vurgu ve duraksamaların da sık kullanımı sonucunda “Samanyolu” şarkısına ait müziğin sözleri, melodiye uygun şekilde oluşturulmuş ve uygulamanın tekdüze bir şekilde şarkı söyleyebilmesi sağlanmıştır.



Şekil 3.18 : Uygulamada XML’den okuma seçeneği

Yöntemlerin karşılaştırılması

Çift-ses ekleme, hece ekleme, farklı ses uzunluklu ve farklı genlik değerleri ile ekleme yöntemleri anlatılmıştır. Her yöntem ayrı ayrı denenerek anlaşılabilirlik açısından istatistiksel olarak sonuçları çıkarılmıştır. Bununla ilgili olarak yapılan çalışma ayrıntılı bir şekilde dördüncü bölümde anlatılacaktır. Şekil 3.19’da “halk” kelimesine ait, her bir yöntem ile oluşturulmuş seslerin dalga şekilleri gösterilmektedir.



Şekil 3.19 : “halk” kelimesinin tüm yöntemler kullanılarak oluşturulan ses dalga şekilleri

3.3.4 Gerçekleme ortamı

Türkçe Metin Sentezleme yazılımı gerçekleştirirken şu araçlar kullanılmıştır.

- Eclipse SpringSource Tool Suite (STS)

Kurumsal Java uygulamaları yazmak için kullanılan Eclipse tabanlı, ücretsiz ve açık kaynak kodlu geliştirme ortamıdır[37]. Uygulamanın ana çatısını oluşturan metin işleme, ses işleme bileşenleri ve bu iki bileşenin kullanılması ile oluşturulan Türkçe Metin Seslendirme Sistemi arayüzü, bu geliştirme ortamı ile geliştirilmiştir. Ayrıca görme engelliler için hazırlanan “Görme engelliler için metin derleyici program” da bu araç ile geliştirilmiştir.

- Visual C++

Microsoft tarafından ilk olarak 1993 yılında yayımlanmış ve C programlama dili tabanlı ticari yazılım geliştirme platformudur (IDE). Windows ile tam uyumlu çalışmaktadır, çünkü Windows’un kendisi de C++ ile yazılmıştır.

Türkçe Metin Seslendirme çalışmasında Mbrola kütüphanelerinin Java programlama dili üzerinden de kullanılabilmesi için kullanılmıştır. Bölüm 3.3.2.3’de de anlatıldığı üzere MbrDll projesi kapsamında Mbrola’ya ait kütüphane dosyaları (dll), Java dilinin anlayacağı dll’lere dönüştürülmüştür.

- Visual Basic

Visual Basic, Microsoft tarafından, Basic (Beginners-All Purpose Symbolic Instruction Code) programlama dili üzerinde geliştirilmiş, olay yönlendirmeli, üst seviye, nesne tabanlı ve görsel bir programlama dilidir. Önceki yıllarda kullanılan temel programlama dillerindeki bir düzenleyici (editor) ekran üzerine satır satır yazılan programlama tekniği yerine, grafiksel kullanıcı arabiriminin (GUI) ve bu arabirim içerisinde program geliştirme amacıyla bulunan entegre geliştirme ortamı (IDE) almıştır. Microsoft Windows ile tam uyumlu çalışmaktadır. Windows bileşenlerine (API, COM, DLL gibi) erişmek çok kolaydır.

Türkçe Metin Seslendirme çalışmasında önce Visual Basic kullanıldı fakat görsel anlamda oldukça kolay yazılım yapılabilmesine rağmen, nesneye yönelik programlama tekniği ihtiyacı nedeniyle, sonrasında Java tercih edilmiştir.

- Standart Widget Tool (SWT)

SWT swing’e bir alternatif olarak IBM tarafından geliştirilen bir gösterim sistemidir. Üzerinde çalıştığı işletim sisteminin grafik kütüphanesini kullandığı için, kullanıcı açısından daha anlaşılır ve tanıdık ekranlar üretilmektedir ve çoğu yerde daha hızlıdır. SWT kütüphanesi, Swing gibi Java içinde gelmemektedir, ayrıca dağıtılmaktadır. En bilinen SWT uygulamaları ünlü java yazılım geliştirme aracı olan Eclipse ve Bittorrent paylaşım programı olan Azureus’tur.

Türkçe Metin Seslendirme çalışmasında görsel arayüzler SWT kullanılarak gerçekleştirilmiştir.

- Goldwave

Goldwave Inc. firması tarafından ilk sürümü Nisan 1993 yılında çıkarılmış, ticari, ses dosyaları işleme ve düzenleme programıdır.

Türkçe Metin Soslendirme çalışmasında, üretilen ve kaydedilen ses dosyalarının karakteristik özelliklerini görsel olarak inceleme, birleştirme kurallarının çıkarılması, gürültünün önlenilmesi ve frekans değışikliğı ile vurgunun sağlanabilmesi gibi çalışmalar Goldwave ve Audacity programları aracılığıyla yapılmıştır.

- Audacity

Audacity, ücretsiz, ses kayıt ve düzenleme için kullanılan açık kaynak kodlu ve günümüzde oldukça popüler bir yazılımdır. Mac OS X, Microsoft Windows, GNU/Linux ve diğer birçok işletim sistemleriyle uyumlu bir şekilde çalışabilmektedir[40].

Türkçe Metin Soslendirme çalışmasında, üretilen ve kaydedilen ses dosyalarının karakteristik özelliklerini görsel olarak inceleme, birleştirme kurallarının çıkarılması, gürültünün önlenilmesi ve frekans değışikliğı ile vurgunun sağlanabilmesi gibi çalışmalar Goldwave ve Audacity programları aracılığıyla yapılmıştır.

- HSQLDB

HSQLDB tamamen Java ile yazılmış küçük boyutlu, açık kaynak kodlu ve ücretsiz bir veri tabanı yönetim sistemidir. Küçük ve kullanım kolaylığı sayesinde her türlü Java proje geliştirim ve test aşamasında tercih edilebilir bir seçenek olarak göze çarpmaktadır. Tüm veritabanı yönetim sistemini kullanabilmek için sadece hsqldb.jar kütüphanesi gerekmektedir[41]. Tüm veritabanı işlemleri bellekten yürütüldüğü için küçük projelerde oldukça hızlı sonuç üretebilmektedir.

Türkçe Metin Soslendirme çalışmasında HSQLDB, ses veritabanının yönetilmesinde kullanılmıştır. Harf-ses dönüşümlerinde kullanılacak harf, ses, sampa karakteri, dosya ismi karakteri gibi bilgiler ile ses veritabanlarına ait dosyaların bulunduğu dizin yolu, veritabanı açıklamaları gibi bilgiler belleğe bu veritabanı aracılığıyla yüklenebilmektedir.

3.3.5 Yazılımın diğer yetenekleri

Gerçeklenen Türkçe Metin Soslendirme yazılımda ayrıca aşağıdaki yetenekler de bulunmaktadır. Uygulamaya erişildiğinde, “Dosya” ana adımı

altında Şekil 3.20’de gösterildiği üzere “Yeni, Kaydet, Aç, Kapat, Yazdır ve Çıkış” adımları listelenmektedir. Bu alt adımların işlevi sırasıyla aşağıda özetlenmiştir.

Yeni : Tüm ekranları temizler ve yeni dosya ile çalışılmasını sağlar.

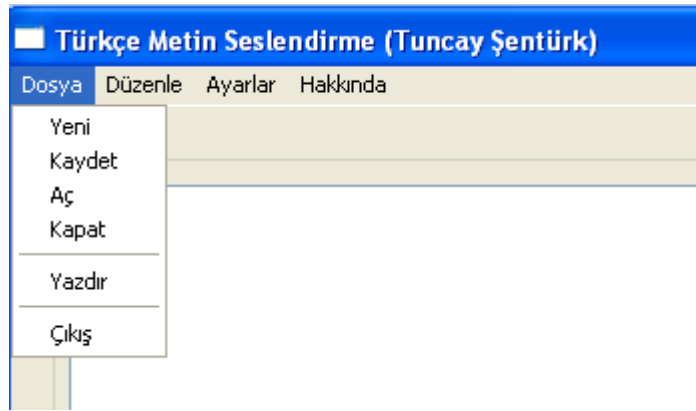
Kaydet : Ekranda yazılı olan metin veya XML’in dosya sistemine kaydedilmesini sağlar. Düz metin seçili ise metin formatında kaydedilebilirken, XML seçili ise “.tms” uzantılı olacak şekilde kaydedilebilmektedir.

Aç : Ekran, daha önceden kaydedilmiş dosya içeriğinin gösterilmesini sağlar. Bu dosya, düz metin veya “.tms” uzantılı dosya olabilir.

Kapat : O an ekranda gösterilen metne ait kullanılan sistem kaynaklarının geri verilmesini ve kullanılmak üzere yeni sayfa açılmasını sağlar.

Yazdır : Ekranda görüntülenen bilgilerin yazıcıya yönlendirilmesini sağlar.

Çıkış : Programın kapatılarak sistem kaynaklarını geri vermesi sağlanır.



Şekil 3.20 : Dosya menüsü adımları

“Düzenle” sekmesinin altındaki adımlar Şekil 3.21’de listelenmiştir.

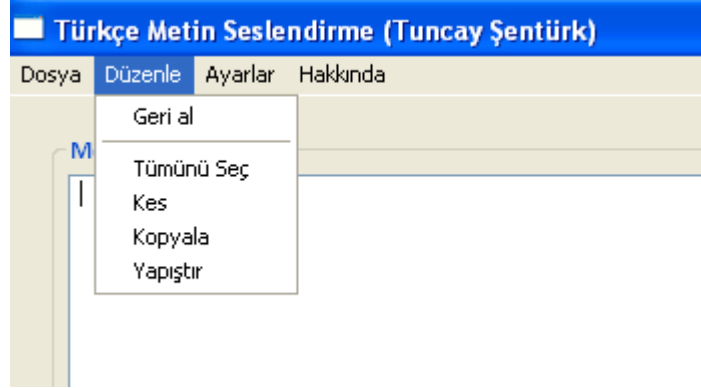
Geri Al : Son yapılan işlemin geri alınmasını sağlar.

Tümünü Seç : Metin kutusunda bulunan tüm metnin seçili hale gelmesini sağlar.

Kes : Seçili bulunan metnin silinip, tekrar kullanılma ihtimaline karşın bellekte (clipboard) tutulmasını sağlar.

Kopyala : Seçili olan metnin belleğe (clipboard) kaydedilmesini sağlar.

Yapıştır : Belleğe (clipboard) kopyalanmış son metnin, imlecin bulunduğu yere kopyalanmasını sağlar.



Şekil 3.21 : Düzenle menüsü adımları

3.3.6 Yazılımın çalışabilmesi için gerekli donanım

Gerçeklenen yazılımın çalışabilmesi için gerekli olan donanım aşağıdaki maddelerle özetlenebilir.

- Windows işletim sistemi
- 150Kbyte boş bellek alanı
- Her ses veritabanı için 50Mbyte boş sabit disk alanı
- Ekran kartı
- Ses kartı
- Hoparlör

4. TÜRKÇE METİN SESLENDİRME UYGULAMASININ DEĞERLENDİRİLMESİ

Türkçe Metin Seslendirme çalışmasında ana hatlarıyla çift-ses eklemeli ve hece eklemeli olmak üzere iki yöntem denenmiştir. Çift-ses ekleme yöntemine göre daha fazla ses kaydı ve bununla ters orantılı şekilde daha az ekleme noktası gerektiren hece ekleme yönteminin daha doğal ve anlaşılır olması için alt çalışmalar gerçekleştirilmiştir. Farklı ses uzunluklu hecelerin eklenmesi yönteminde her sesin farklı uzunlukta oluşturulabilmesi yoluna gidilmiş iken, farklı genlik değerleri ile ekleme yönteminde ise hecelere ve dolayısıyla kelimelere vurgu verilmeye çalışılmıştır.

Her bir yöntem için değerlendirmenin yapılabilmesi amacıyla 10 cümle belirlenmiştir. Her bir yöntem ile üretilen sesler farklı yaş gruplarındaki altı kişiye farklı zamanlarda dinletilmiştir. Dinleyicilerden gelen cevaplar üzerine anlaşılabilirlik hesabı yapılmıştır.

4.1 Anlaşılabilirliğin hesaplanması

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 4.1'de listelenmiştir. Dinleyiciler, dinledikleri sesleri değerlendirerek beş seviyeden birini uygun görürler[20].

Çizelge 4.1 : 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 sonuç 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 anlaşılabilirliğin hesaplanmasına dahil olması gereken parametreler aşağıda özetlenmiştir.

doğru harf sayısı : dinleyicinin anladığı cümle ve gerçek cümledeki çakışan harf sayısını gösterir. (Anlaşılabilirlik 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, anlaşılabilirlik 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 kaçıcı denemenin olduğunu gösterir. (Cümle ilk 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 **(4.1)** anlaşılabilirlik formülü elde edilmiştir.

$$p(x) = \sum_{n=1}^3 \left(\frac{\text{doğru harf sayısı} - \text{önceki doğru harf sayısı}}{\text{toplam harf sayısı}} \cdot \frac{1}{n} \right) \quad (4.1)$$

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.

4.2 Cümlelerin belirlenmesi

Test sonucunun daha anlamlı olması için, (3 ve 7 yaşında dinleyicilerin de olduğu düşünülerek), her yaştaki insana hitap edecek, net cümleler seçilmeye gayret edilmiştir. Seçilen bu cümleler Çizelge 4.2’de gösterilmektedir.

Çizelge 4.2 : Dinleyicilere dinletilen 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

4.2.1 Cümlelerin dinleyicilere dinletilmesi

Belirlenen her cümle dinleyicilere aynı ortamda ve farklı zamanlarda dinletilmiştir. Tüm yöntemlerin test edilebilmesi amacıyla, daha önceden dinletilmiş yönteme ait cümlelerin hatırlanmaması için bir ay gibi bir sürenin beklenmesi öngörülmüştür.

Cümlelerin hatırlanmaması için bir ay gibi bir 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.

4.2.2 Sonuçların değerlendirilmesi

Altı farklı yaş grubundaki dinleyicilerle yapılan deney neticeleri, formülden geçirildikten sonra Çizelge 4.3'te oluşturulan not dağılımı ortaya çıkmıştır. Her bir dinleyicinin, tüm cümleler için vermiş olduğu cevaplara ilişkin yüzde cinsinden notu, yine her bir yöntem için belirtilmiştir. Ayrıca her bir yöntem için ayrı ayrı cümlelerin ara ortalaması da hesaplanmış ve en sonunda da her bir yöntemin almış olduğu genel ortalama not değeri Çizelge 4.4'deki gibi bulunmuştur. Çift-ses eklemeli yöntem kullanılarak yapılan deneyde anlaşılabilirlik oranı %91.5 iken bu değer hece eklemeli yöntemde %96.1'e yükselmiştir. Genlik ve ses uzunluğu değişimi ile vurgu çalışmasında ise anlaşılabilirlik %98 olarak ölçülmüştür.

Çizelge 4.3 : Dinleyici cevaplarından oluşturulan not dağılımı

Cümle 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)	86,842	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	97,807	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)	77,631	76,315	77,631
		Arda (7 yaş)	16,228	22,368	77,631
		Bahar (15 yaş)	82,894	92,105	82,894
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	58,771	91,228	98,684
		Mebrure (60 yaş)	82,017	98,684	100
		Ortalama	69,590	80,116	89,473
4	Parkta oynayan çocuklar uçurtma uçurdu	Batu (3 yaş)	66,667	55,882	100
		Arda (7 yaş)	23,529	100	67,156
		Bahar (15 yaş)	77,941	89,7058	100
		Tarık (32 yaş)	95,588	100	100
		Pınar (34 yaş)	98,529	100	100
		Mebrure (60 yaş)	91,176	97,058	100
		Ortalama	75,571	90,441	94,526
5	Akşam yatmadan önce süt içtim	Batu (3 yaş)	83,974	100	100
		Arda (7 yaş)	100	100	100
		Bahar (15 yaş)	100	100	100
		Tarık (32 yaş)	100	100	100
		Pınar (34 yaş)	94,230	100	100
		Mebrure (60 yaş)	98,076	100	100
		Ortalama	96,047	100	100

Çizelge 4.3 (devam) : Dinleyici cevaplarından oluşturulan not dağılımı

6	Televizyonda çizgi film seyretmeyi çok severim	Batu (3 yaş)	68,69918699	93,90243902	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	94,78319783	98,98373984	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	97,91666667	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	94,58333333	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	Dışlerinizi her gün iki kere fırçalamalısınız	Batu (3 yaş)	83,75	84,16666667	92,5
		Arda (7 yaş)	56,25	80,83333333	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	87,91666667	94,16666667	97,29166667

Çizelge 4.4 : Genel ortalamalar

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

Çizelge 4.5’de ise kişi bazında cümlelere verilen cevaplara göre değerlendirme yapılmıştır. Sırasıyla çift-ses eklemeli, hece eklemeli ve vurgulu (ses uzunluğu ve genlik değişimi ile) yöntemler değerlendirildiğinde, anlaşılabilirliğin arttığı açıkça gözlemlenmiştir. Hemen hepsinde yöntemlerde sağa doğru geçildikçe anlaşılabilirlik artmıştır.

4.2.3 Çalışmanın yaptığı katkılar

İncelenmiş olan diğer çalışmalardaki anlaşılabilirliğin artırılması için özetle aşağıdaki çalışmalar gerçekleştirilmiştir.

Ulama : Türkçe Metin Seslendirme sisteminde anlaşılabilirliğin daha da artırılması için Türkçe dilbilimi çatısı altında çalışmalar yapılmıştır. Bunlardan birisi

3.3.3.1 Metin İşleme ve 3.3.3.2 Ses İşleme bölümlerinde de anlatıldığı üzere ulama çalışmasıdır. Bu çalışma ile az da olsa anlaşılabilirliğin ve doğrallığın artması sağlanmıştır.

Ses dosyalarının dengelenmesi : 3.3.2.4 Oluşturulan ses dosyalarının genliklerinin dengelenmesi çalışması ile ekleme noktalarındaki patlamalar engellenmeye çalışılmıştır. Bu doğrultuda kaydedilen tüm ses dosyalarının genlikleri arasında önemli farklar olmaması sağlanmıştır.

Ses uzunluklarının ve genliklerinin değiştirilebilmesi : Türkçedeki tüm heceler, farklı kelimelerde farklı uzunlukta söylenebilmektedir. Kelimenin başında veya sonunda, cümlenin başında veya sonunda olmasına göre farklı uzunluklarda okunabilmektedir. Hecelerin doğrudan eklenmesi ile bu özellik sağlanamazken, bu çalışmada her ses için istenen uzunluk milisaniye cinsinden verilebilmekte ve dolayısıyla üretilen ses daha doğal ve anlaşılır hale gelebilmektedir. Aynı şekilde hecelerin genliklerinin belli kurallar doğrultusunda artırılması ile vurgu sağlanmaya çalışılmış ve anlaşılabilirlik için etkili olmuştur.

Ayrık bileşenler ile esneklik kazandırılması : Metinden XML ve XML’den ses üretme bileşenleri tamamen birbirinden bağımsız çalışabildiği için, yapılacak olan dilbilimsel çalışmalar ile uygulamaya kolay bir şekilde yeni özellikler eklenebilecektir. Gevşek bağlaşım sayesinde her iki bileşen tamamen ayrılmış ve üstüne görme engelliler için düzenleyici uygulaması kolaylıkla eklenebilmiştir.

Çizelge 4.5 : Dinleyici cevaplarına göre oluşturulan başarı oranları

	Batu (Yaş : 3)			Arda (Yaş : 7)			Bahar (Yaş : 15)			Tarık (Yaş : 32)			Pınar (Yaş : 34)			Mebrure (Yaş : 60)		
Bu sabah erken kalktım	86,84	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Bu akşam çok yemek yedim	100	100	100	100	100	100	92,5	100	100	100	100	100	100	100	100	100	100	100
Çok fazla kar yağdığı için annem işe gidemedi	77,63	76,32	77,63	16,23	22,37	77,63	82,89	92,11	82,89	100	100	100	100	100	100	82,02	98,68	100
Parkta oynayan çocuklar uçurtma uçurdu	66,67	55,88	100	23,53	100	67,16	77,94	89,71	100	95,59	100	100	100	100	100	91,18	97,06	100
Akşam yatmadan önce süt içirim	83,97	100	100	100	100	100	100	100	100	100	100	100	100	100	100	98,08	100	100
Televizyonda çizgi film seyretmeyi çok severim	68,7	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Babamla futbol oynadık	100	87,5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Artık yatma vakti geldi	100	100	100	87,5	100	100	90	100	100	100	100	100	100	100	100	90	100	100
Polis hırsızı yakaladı	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Dişlerinizi her gün iki kere fırçalamalısınız	83,75	84,17	92,5	56,25	80,83	91,25	91,25	100	100	100	100	100	100	100	100	96,25	100	100
ORTALAMA	86,76	90,39	97,01	78,35	90,32	93,6	93,46	98,18	98,29	99,56	100	100	100	100	100	95,75	99,57	100

Görme engelliler için düzenleyici uygulaması : Türkçe metin seslendirme uygulamasının yanısıra, görme engelliler için özel metin yazma/düzenle uygulaması geliştirilmiştir ve kullanımlarına sunulmuştur. Görme engelliler için metin düzenleyici uygulaması Bölüm 4.3’te detaylı olarak anlatılmaktadır.

4.2.4 Görme engelliler için metin düzenleyici program

Türkçe metin seslendirme sisteminin her iki bileşeni de kullanılarak görme engelliler için metin düzenleyici program 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” kelimesinin yazımı sırasında, tuştakımından tuşlanan “M”, “E”, “R”, “A” ve “K” harfleri sırasıyla “ME”, “E”, “RE”, “A”, ve “KE” şeklinde seslendirilmektedir.

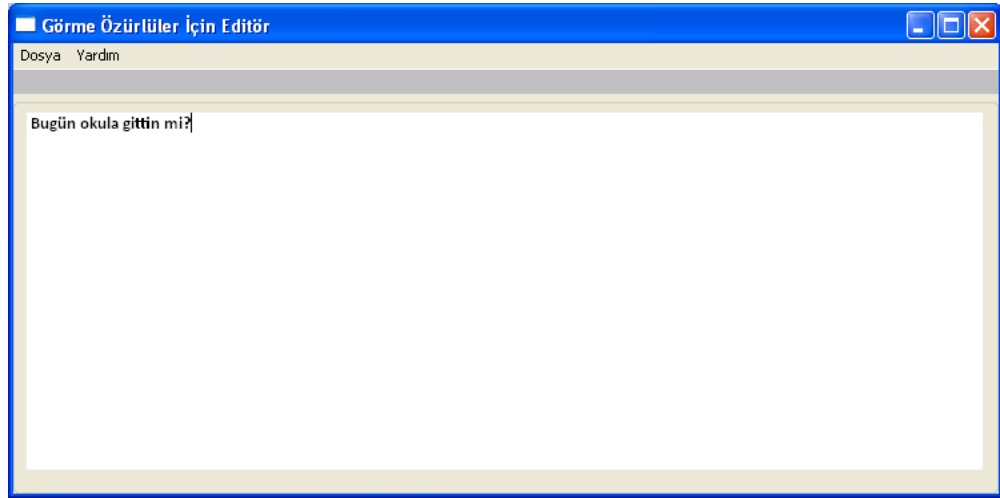
Rakamlar tuşlandıkça seslendirilmesi sağlanmıştır. Örneğin tuştakımından girilen “12” sayısı için sırasıyla “BİR” ve “İKİ” seslendirmesi yapılmaktadır.

Noktalama işaretleri tuşlandıkça, Çizelge 4.6’da tanımlandığı şekliyle seslendirilme yapılmaktadır.

Çizelge 4.6 : Noktalama işaretlerinin seslendirilmesi

Noktalama işareti	Seslendirme
,	VİRGÜL
;	NOKTALI VİRGÜL
.	NOKTA
:	İKİ NOKTA ÜST ÜSTE
?	SORU İŞARETİ
!	ÜNLEM
"	TIRNAK
'	TEK TIRNAK
\r veya \n	SATIR SONU

- 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 4.1).



Şekil 4.1 : Görme engelliler için metin düzenleyici programı

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 bildirilmektedir.
- (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 kelimenin seslendirilmesi sağlanmaktadır.

5. SONUÇ VE ÖNERİLER

Bilgisayarla insan ve makine arasındaki sözel iletişim, son yıllarda önemi gittikçe artan bir konudur. Dünyada bu alanda uzun süredir yapılan çalışmalar sonucu, anlaşılabilirliği 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 5.1’deki gibi girilen metnin tüm yöntemler doğrultusunda, ses üretebilmesi sağlanmıştır.

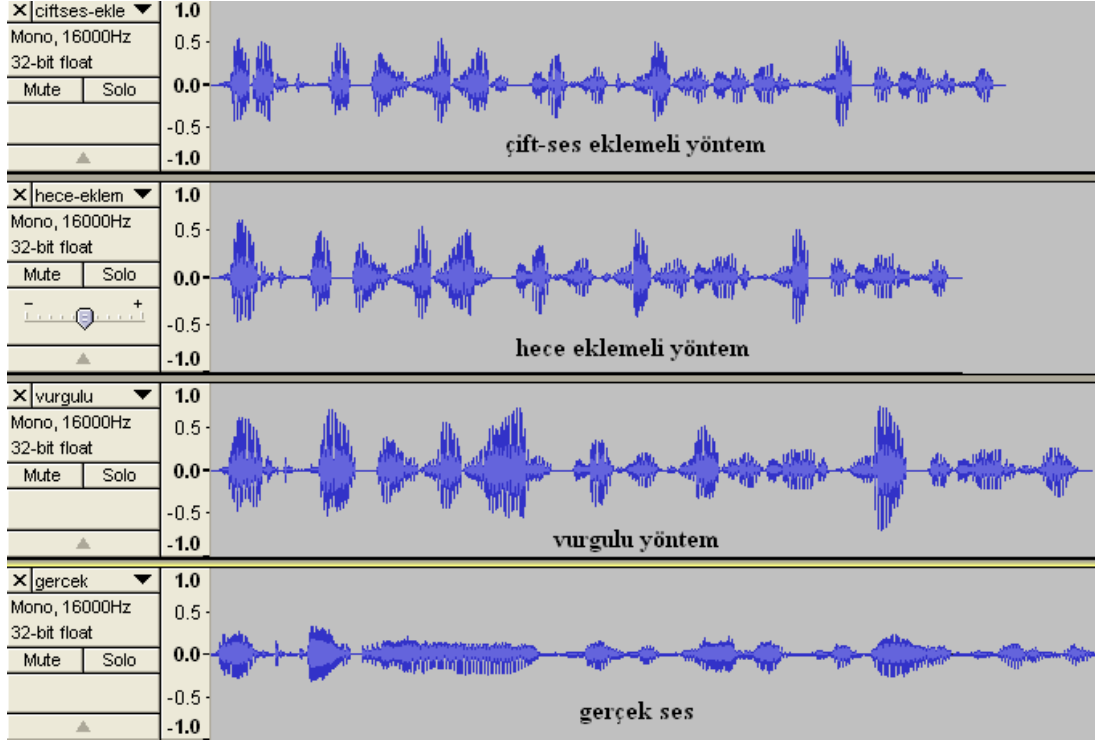
Türkçedeki ikisi dış kaynaklı olan sekiz farklı hece tipi üzerinde çeşitli incelemeler yapılmıştır. Her bir hece tipinin ekleme yöntemleri üzerinde formüller gerçekleştirilmiştir.

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 anlaşılabilirliği 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.



Şekil 5.1 : “Parkta oynayan çocuklar uçurtma uçurdu” cümlesinin farklı yöntemlerle oluşturulmuş ses dalga şekilleri

Karşılaşılan bir diğer sorun da, hecelerin, sözcüğün içindeki konumlarına bağlı olarak değişiklik göstermesidir. Yani, eğer bir heceye ait ses dosyası, sözcük başından çıkartılarak elde edilmiş ise, aynı ses dosyasının, sözcük sonunda veya ortasında kullanılması istenmeyen vurgulara ve bozulmalara yol açabilmektedir. Bu yüzden ses dosyaları mümkün olduğunca kelimeler içinden seçilerek değil de tamamen tekdüze ve ayrı bir şekilde oluşturulmalıdır. Bu da vurgu ve tonlamayı oldukça zorlaştırmaktadır.

Türk alfabesindeki 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 fonetik alfabesi çıkartılmıştır ve “lale” gibi sözcüklerde hangi sesin kullanılacağını, aykırı kelimeler sözlüğünden elde edilmesi hedeflenmiştir.

Sonu olarak, alıřmadan daha doęal sesler elde etmek iin frekans alanı zerinde alıřma yapılmalıdır. Pitch deęerleri zerinde durularak ve doęal dil iřleme desteęinin de alınması ile ok doęal sonuların elde edilebileceęi alıřma sonucunda ispatlanmıřtır.

KAYNAKLAR

- [1] Rıfat YAZICI ve Cemal KÖSE, "Kısıtlamasız Türkçe Ses Sentezi" Sinyal İşleme Uygulamaları (SİU) Konferansı, 1992
- [2] Juergen Schroeter, *Text to-Speech (TTS) Synthesis*, AT&T Laboratories, 2005
- [3] Flanagan J., *Speech Analysis, Synthesis, and Perception*, Springer-Verlag, Berlin-Heidelberg-New York, 1972
- [4] Schroeder M., *A Brief History of Synthetic Speech*, Speech Communication vol.13 pp.231-237, 1993
- [5] Güldalı K., Türkçe Metin Seslendirme, İstanbul Teknik Üniversitesi, 2009
- [6] <<http://www.haskins.yale.edu/featured/heads/SIMULACRA/kempelen.html>>, alındığı tarih 20.02.2010
- [7] Klatt D., Review of Text-to-Speech Conversion for English, Journal of the Acoustical Society of America, JASA vol.82(3)737-793,1987
- [8] <<http://www.haskins.yale.edu/featured/heads/SIMULACRA/wheatstone.html>>, alındığı tarih 20.02.2009.
- [9] Eker B., Turkish Text To Speech System, 2002
- [10] Dutoit T., "An Introduction to Text to Speech Synthesis", pp 26-32, 1997
- [11] Rosen G., The DAVO articulatory synthesizer, MIT, 0.28, 1958
- [12] Allen, J., Hunnicutt, S., Klatt, D., From Text to Speech: The MITalk System, Cambridge University Press, 1987
- [13] Dutoit T., A Short Introduction to Text-to-Speech Synthesis <http://tcts.fpms.ac.be/synthesis/introts_old.html>, alındığı tarih 25.02.2010
- [14] <<http://www.acapela-group.com>> , alındığı tarih 25.02.2010
- [15] Ljungqvist M., Lindström A., Gustafson K., A New System for text-to-Speech and Its Applications to Swedish, ICSLP94 (4) : 1779-1782, 1994
- [16] Mönius B., Schroeter J., Santen J., Sproat R., Olive J., Recent Advances Multilingual Text-to-Speech Synthesis, Fortschritte der Akustik, DAGA, 1995
- [17] <<http://www.research.att.com/~ttsweb/tts/index.php>> , alındığı tarih 26.02.2010
- [18] <<http://www.nuance.com/realSpeak>> , alındığı tarih 26.02.2010
- [19] Festival Project Homepage <<http://www.cstr.ed.ac.uk/projects/festival>>, alındığı tarih 26.02.2010
- [20] Lemmetty S., Review of Speech Synthesis Technology, Helsinki University of Technology, 1999

- [21] **MBROLA Project Homepage** <<http://tcts.fpms.ac.be/synthesis/mbrola.html>>, alındığı tarih 26.02.2010
- [22] **Sestek Anasayfası** <<http://www.sestek.com.tr/gvztts.html>>, alındığı tarih 01.03.2010
- [23] **Adalı E.**, Doğal Dil İşleme, 2010
- [24] **Türkçe İmla Kılavuzu** - Türk Dil Kurumu, 2000
- [25] **Prof. Dr. Muharrem ERGİN**, "Üniversiteler İçin Türk Dili", Bayrak Basım/Yayın/Tanıtım, İstanbul, 1992
- [26] **Flanagan J. L., Allen J. B., M. A. Hasegawa-Johnson**, "Speech Analysis Synthesis and Perception", ch6, ch2., 2008
- [27] **Güner L. Ergenç D.**, Sesin Doğası ve Oluşumu <<http://www.jandarma.tsk.tr/kriminal/turkish%20internet/anasayfa/bil arinde dosyalar/yazilar dosyalar/bilarinde5.pdf>>, alındığı tarih 03.03.2010
- [28] **Ergenç İ.**, *Konuşma Dili ve Türkçenin Söyleyiş Sözlüğü*, 1995
- [29] **Kröger B.**, Minimal Rules for Articulatory Speech Synthesis, Proceedings of EUSIPCO92(1):331-334, 1992
- [30] **Rahim M., Goodyear C., Klejin B., Schroeter J., Sondhi M.**, *On the Use of Neural Networks in Articulatory Speech Synthesis*, Journal of the Acoustical Society of America, JASA vol.93(2):1109-1221,1993
- [31] **Ertaş F., Eskidere Ö.**, *Yazılım Tabanlı Sözcük Sentezleyici*, DEÜ Fen ve Mühendislik Dergisi Cilt:3 Sayı:1 sh.1-27, 2001
- [32] **Klatt, D. H.**, "Review of text-to-speech conversion for English", MIT, 1987, <<http://americanhistory.si.edu/archives/speechsynthesis/dk737a.htm>>, alındığı tarih 01.03.2010
- [33] **Jurafsky and Martin**, *Speech and Language Processing*, 2ed (Draft), 2000
- [34] "Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones" *Speech Communication* (Elsevier Science Publishers, Amsterdam, Netherlands), vol. 9, Nos. 5/6, pp. 453-467, 1990
- [35] **Wave Dosya Formatı** <<http://ccrma.stanford.edu/courses/422/projects/WaveFormat>>, alındığı tarih 01.03.2010
- [36] **SAMPA Türkçe** <<http://www.phon.ucl.ac.uk/home/sampa/turkish.htm>>, alındığı tarih 07.03.2010
- [37] **SpringSource Tool Suite Homepage**, <<http://www.springsource.com/products/sts>>, alındığı tarih 17.03.2010
- [38] **SWT Homepage** <<http://www.eclipse.org/swt>>, alındığı tarih 17.03.2010
- [39] **How to Call Native (DLL) Code From Java Using JNI** <<http://support.microsoft.com/kb/222092>>, alındığı tarih 15.11.2009
- [40] **Audacity Homepage** <<http://audacity.sourceforge.net>>, alındığı tarih 17.03.2010

[41] **HSQLDB Homepage** <<http://www.hsqldb.org>>, alındığı tarih 17.03.2010

ÖZGEÇMİŞ



Ad Soyad: Tuncay ŞENTÜRK

Doğum Yeri ve Tarihi: Of / 1978

Adres: Gardenya 5-1B Daire 32 Ataşehir / İSTANBUL

Lisans Üniversitesi: İstanbul Teknik Üniversitesi
Kontrol ve Bilgisayar Mühendisliği