

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKÇE METİN SESLENDİRME

**YÜKSEK LİSANS TEZİ
Kenan GÜLDALI**

Anabilim Dalı : Bilgisayar Mühendisliği

Programı : Bilgisayar Mühendisliği

EKİM 2009

TÜRKÇE METİN SESLENDİRME

**YÜKSEK LİSANS TEZİ
Kenan GÜLDALI
504061536**

Tezin Enstitüye Verildiği Tarih : 07 Eylül 2009

Tezin Savunulduğu Tarih : 13 Ekim 2009

**Tez Danışmanı : Prof. Dr. Eşref ADALI (İTÜ)
Diğer Jüri Üyeleri : Yrd. Doç. Dr. Gülşen CEBİROĞLU
ERYİĞİT (İTÜ)
Yrd. Doç. Dr. Banu DİRİ (YTÜ)**

EKİM 2009

Aileme,

ÖNSÖZ

Öğrenim hayatımda ve çalışmalarımnda desteklerini esirgemeyen aileme ve tez danışmanım Sn. Prof.Dr. Eşref ADALI'ya teşekkürü bir borç bilirim.

Eylül 2009

Kenan GÜLDALI

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	v
İÇİNDEKİLER.....	vii
KISALTMALAR	ix
ÇİZELGE LİSTESİ.....	xi
ŞEKİL LİSTESİ	xiii
ÖZET	xv
SUMMARY.....	xvii
1. GİRİŞ	1
1.1 Amaç.....	2
1.2 Ses Sentezleme Sistemlerinin Gelişimi ve Tarihçesi	3
1.3 Önemli Metinden Ses Üretme Sistemleri	4
1.3.1 MITalk.....	5
1.3.2 Infovox.....	5
1.3.3 AT&T Labs Natural Voices & Bell Labs TTS	5
1.3.4 ETI Eloquence / Speechworks / Nuance	5
1.3.5 MBROLA	6
1.3.6 Festival TTS.....	6
1.3.7 GVZ TTS	6
2. SES BİLİMİ VE İNSANDA SES ÜRETİMİ	7
2.1 Türkçe’de Sesler.....	7
2.2 Sesin Sayısal Kodlanması.....	8
2.3 İnsanda Ses Üretimi.....	9
2.4 Ses Sentezleme Teknikleri.....	11
2.4.1 Biçimlendirici (Formant) sentezleyiciler.....	13
2.4.2 Mafsallı ya da telaffuz (Articulatory) sentezleyiciler.....	15
2.4.3 Bağlamalı (Concatenative) sentezleyiciler	15
3. TÜRKÇE İÇİN METİN SESLENDİRME SİSTEMİ	17
3.1 Problemler.....	17
3.1.1 Metinden sesçil birimlere dönüştürme	17
3.1.1.1 Metin önışleme	17
3.1.1.2 Doğru bir söyleyiş için dilbilimsel veri çıkarımı	18
3.1.1.3 Ölçü analizi	18
3.1.2 Sentezleme aşaması problemleri.....	19
3.2 Türkçe Metin Seslendirme Sistemi	20
3.2.1 Ses dosyaları	20
3.2.2 Veri tabanı hazırlanması	22
4. GERÇEKLEME	25
4.1 Birleştirme ve Veri Hazırlama Kurallarının Oluşturulması.....	25
4.1.1 Hece veri tabanının oluşturulması.....	25
4.1.2 Birleştirme Kuralları.....	27
4.2 Yazılım Gerçekleme.....	30
5. SONUÇ VE ÖNERİLER	37

KAYNAKLAR..... 38

KISALTMALAR

LPC	: Linear Predictive Coding
KHz	: Kilo Hertz
TDK	: Türk Dil Kurumu

ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1 : Türkçede Sesli Harfler	7
Çizelge 2.2 : Türkçede Sessiz Harfler	7
Çizelge 2.3 : Oluşum Şekillerine Göre Türkçe Sesler.....	8
Çizelge 2.4 : Ertaş F., Eskidere Ö. Sözcük Birimlerinin Seçimi Çizelgesi	12
Çizelge 3.1 : RIFF Veri Bölgesi.....	20
Çizelge 3.2 : FORMAT Veri Bölgesi.....	21
Çizelge 3.3 : DATA Veri Bölgesi	21
Çizelge 3.4 : Türkçe İçin Hece Sayıları.....	23
Çizelge 4.1 : Türkçeye özgü karakterlerin UTF-8 karşılıkları.....	31
Çizelge 5.1 : Dinleyici yorumlarından oluşturulmuş ortalamalar	37

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1 : Metinden konuşma sentezlemenin temel aşamaları.	2
Şekil 1.2 : Kratzenstein'in ses biçimlendirme cihazları.	3
Şekil 1.3 : Wheatstone'un geliştirdiği Kempelen ses cihazı.....	4
Şekil 2.1 : Ses Sinyalinin Örneklenmesi	8
Şekil 2.2 : İnsan Ses Yolunun Yapısı	9
Şekil 2.3 : İnsan Ses Üretim Organları	10
Şekil 2.4 : Kaskat Biçimlendiriciler	13
Şekil 2.5 : Paralel Biçimlendiriciler	14
Şekil 3.1 : Ölçülü Söyleyişi Etkileyen Faktörler.....	19
Şekil 3.2 : Ötümsüz ve Ötümlü Ses.....	22
Şekil 4.1 : İşaretleme işlemi.....	26
Şekil 4.2 : Hece birleşme noktasında boğumlar.....	29
Şekil 4.3 : Ulaştırma kelimesinin veri tabanından alınmış gerçek dalga formu ve hecelerle oluşturulmuş dalga formu	29
Şekil 4.4 : Acıkmak kelimesinin kesilmiş hecelerle işlem görmüş ve doğrudan birleştirilmiş dalga formu.....	30
Şekil 4.5 : Birleştirme noktasına sessizlik ekleme	32
Şekil 4.6 : Gerçeklenen Hece Birleştirmeli Sistemde Hece Uzatma İşlemi	33
Şekil 4.7 : Gerçeklenen Hece Birleştirmeli Sistem	34
Şekil 4.8 : Gerçeklenen Hece Birleştirmeli Sistemin Kullanıcı Arayüzü	35

TÜRKÇE METİN SESLENDİRME

ÖZET

Bu tez çalışmasında ses birimi uzunluğu olarak heceyi kullanan, bağlamalı sentezleme tekniğine dayanan bir metin seslendirme sistemi gerçekleştirilmiştir. Bir metin sentezleyici gerçekleştirirken, oluşturulan konuşma çıkışının kalitesinin yüksek olması, hafıza gereksiniminin düşük olması, uygulamanın karmaşıklığının çok fazla olmaması, dolayısıyla hesaplama hızının yüksek olması amaçlanır.

Tasarlanan sistemde kayıtlı kelimelerden hecelerin kesilmesi ve bu hecelerin birleştirilmesi ile ilgili kurallar oluşturulmuştur. Bu kurallar kelimelerde hecelerin birleştirme yerlerindeki seslerin oluşum şekillerine(nefesli, nefessiz, patlamalı vb.) ya da üretim yerlerine bağlı olarak oluşturulmuştur. Kurallar oluşturulurken Türk Dil Kurumu sesli sözlük veri tabanının ses dosyaları üzerinde birçok deneme yapılmıştır. Bu denemeler ışığında elde edilen kurallar daha sonra birleştirme yazılımına eklenmiştir. Gerçeklenen bu sözcük sentezleyici yazılımı için TDK veri tabanından küçük bir hece veri tabanı oluşturulup denenmiştir. Veri tabanının oluşturulması aşaması sistemin kalitesini son derece etkilemektedir. Veri tabanı hazırlanırken, hecelere ait ses dosyalarına ek olarak birleştirme işleminde sisteme gerekli olan parametreler de oluşturulur. Her hece için oluşturulan bu parametreler bir dosyada tutulur ve uygulamada girdi olarak kullanılır. Sistem, girişte aldığı metin bilgisinden heceleri oluşturur. Daha sonra bu hecelere ait parametre dosyalarını kullanarak heceleri kayıtlı oldukları dosyalardan alır ve bunları birleştirmeye başlar. Bu aşamada hecelerin birleşme noktalarında seslerin türlerine göre belirlenen kurallar uygulanır ve gerçek ses dosyalarındaki dalga şekillerine benzer doğruluk oluşturulmaya çalışılır. Bu doğruluk gerekli yerlerde hecelerin başında ya da sonunda uzatma ve boğumlama yapılarak sağlanmaya çalışılmıştır.

Oluşturulan veri tabanı kullanılarak yapılan seslendirmelerde, heceleri doğrudan birleştirmeye göre anlaşılabilirlik açısından daha iyi sonuçlar elde edilmiştir.

TURKISH TEXT TO SPEECH SYSTEM

SUMMARY

In this MS thesis, a text to speech system which uses syllable as sound unit length and depends on concatenative synthesis method was developed. When developing a text to speech system, it should be aimed to have a high quality speech output, with small memory requirement using an application which is not complex. Consequently computation speed should be high.

In this project, rules for cutting syllables from word database and concatenating these syllables were created. These rules were formed according to the formation and place of articulation of sound (e.g. voiced, unvoiced, plosive). While forming the rules, many tests were done on the sound files of the voice dictionary database of Turkish Language Association. The rules which were achieved as the result of these tests were later added to the concatenation software. For using with the developed word synthesizer software, a small syllable database was formed from Turkish Language Association database. Database preparation is very important for the quality of text to speech system. While preparing the database, parameters needed for the concatenation process are prepared in addition to the sound files related to the syllables. These parameters are formed separately for each syllable and are saved in a file for each syllable which is used as input of the application. The system forms the syllables from the text information received from the input. Consequently, it takes the syllables from the related files, using the parameters for each syllable and starts to synthesize them. At this step, the rules which were developed according to the sound types at the concatenation points of syllables are used and natural sounds which are similar to the waveforms in real sound files are tried to be developed. It is tried to have this naturalness by putting extension or articulation at the beginning or end of the syllable in the required places.

As the result of text to speech trials using the prepared database, better results were achieved in terms of understandability compared to direct concatenation of syllables.

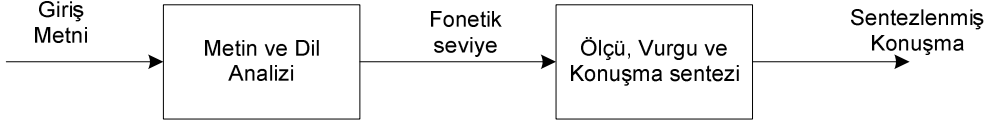
1. GİRİŞ

Teknolojinin gelişmesi ile insan-makine iletişimi gelişmiş ve çeşitlenmiştir. Bu iletişime getireceği kolaylıklardan dolayı ses uygulamaları şüphesiz önemli bir yere sahip olmuştur. Ses tanıma, konuşmacı doğrulama ve sesli komut işleme gibi uygulamalarda ses, makine için bir giriş parametresi olurken, metinden ses sentezleme işleminde bir çıkış parametresidir. Günümüze kadar metinden ses(konuşma) elde etmek için ticari ya da akademik birçok çalışma yapılmış ve birçok yöntem geliştirilmiştir. Bu çalışmaların çoğunda istenen doğruluk yakalanamadığı için halen geliştirilmeye devam etmektedir. İlk başlarda sadece mekanik çözümler ile başlayan çalışmalar günümüzde bilgisayar teknolojisinin gelişmesiyle farklı yöntemlerin denenebilmesinin önünü açmıştır. Bu yöntemlerin incelenmesi Türkçe için yeni bir sistem önerilmesinde faydalı olmaktadır.

Metinden yapay konuşma sinyallerinin üretilmesi işleminin kabul edilebilir kalitede sonuçlar vermesi için üzerinde çalışılan dilin ses özelliklerinin modellenmesi büyük önem taşımaktadır. Dilin ses özellikleri kullanılacak olan yöntemlerin seçilmesinde birincil önceliklidir. Mevcut sistemler genelde belirli bir dil için tasarlanmış olduğu için yeni bir dil için bu sistemleri uyarlamak zor olmakta ve başarılı sonuçlar vermemektedir.

Metinden konuşma üretme işlemi Şekil 1.1 'de gösterildiği gibi temelde iki aşama ile özetlenebilir: Birinci aşama giriş metninin sentezleme işlemine uygun hale getirilmesi için yapılan metin çözümleme ve dil analizi aşamasından oluşmaktadır. İkinci aşama ise birinci aşamadan elde edilen bilgiler ile gerçekleştirilen konuşma(ses) sentezleme aşamasıdır. Metin çözümleme aşamasında metin çözümlenerek sentezleme aşamasının kullanabileceği şekilde bir çıkış üretilir. Bu çıkış tasarıma uygun parçalanmış birimleri ve bazı ek süre bilgilerini içerir. Bu bilgiler ses sentezleme uygulamasına verilir ve sesin üretilmesi aşaması bu bilgiler kullanılarak gerçekleştirilir. Çözümleme işleminden tahmin edilebilecek olan süre bilgileri doğruluk, tonlama ve vurgu açısından önem arz etmektedir. Bu açıdan

düşünüldüğünde metnin sentezleme aşamasına geçirilişine kadar olan işlemler, yüksek başarılı bir sistem için çok karmaşık olabilmektedir.



Şekil 1.1 : Metinden konuşma sentezlemenin temel aşamaları.

Yapay ses üretmenin en kolay yolu, kelime ya da cümle gibi uzun birimlerin kaydedilmesi ve ard arda çalınmasıdır. Tahmin edilebileceği gibi bu bağlamalı (concatenative) yöntem kaliteli ve çok doğal bir ses üretir, fakat çok küçük bir kelime hazinesine sahiptir. Söylenebilecek kelimelerin sayısı oluşturulabilen veri tabanının büyüklüğü ile orantılıdır. Bu şekilde genelleştirilmiş bir sistem oluşturulamayacağı açıktır. Sadece belirli bir amaca yönelik kısıtlı bir veri tabanı için seslendirme sağlanabilir. Bu nedenle daha genel amaçlı bir sistem için fonem, iki fonem(diphone), üç fonem(triphone) ya da hece gibi daha kısa ses birimleri kullanılmalıdır.

Diğer bir yöntem ise kaynak-filtre modelini temel alan biçimlendirici (formant) sentezleyicilerdir. Bu yöntem bir ses kaynağı, biçimlendirme frekansları ve filtreler ile ses sinyallerinin modellenmesini amaçlamaktadır.

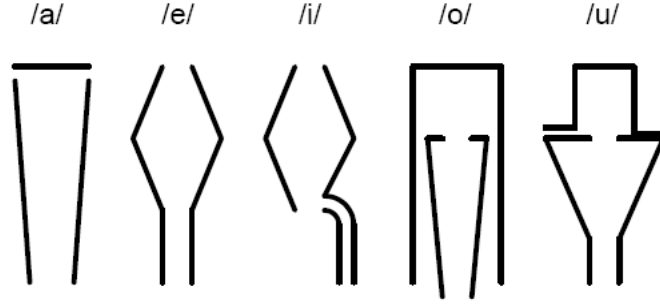
İnsanın ses yolunu modellemeyi amaçlayan mafsallı (articulatory) sentezleyiciler kuramsal olarak en doğru yöntem gibi görülse de karmaşıklığından dolayı gerçekleşmesinin çok zor olduğu bilinmektedir.

1.1 Amaç

Araştırma amaçlı kullanılacak Türkçe için metinden ses üretme sistemleri halen kabul edilebilir kalitede değildir. Göreceli olarak kaliteli olan ticari uygulamalar ise çok büyük veri tabanı ve işlemci gücü istemektedir. Bu tezin amacı, mevcut sistemlerden, araştırma amaçlı kullanılacak olanlardan Türkçe desteği olanların araştırılıp incelenmesi, kullanılan yöntemler ve tasarımların incelenmesi, Türkçe desteği olmayanlar için Türkçeleştirme çalışmasının uygun olup olmadığının araştırılması ve sonuç olarak, bu araştırmalar ışığında Türkçe bir metinden ses sentezleme sistemi tasarlamaktır.

1.2 Ses Sentezleme Sistemlerinin Gelişimi ve Tarihçesi

Ses sentezi sistemleri başlangıçta mekanik çalışmalarla başlamıştır. 1779 yılında Rus profesör Christian Kratzenstein beş tane sesli harfi müzik enstrümanına benzer yapılarla kısmen ses çıkarmayı başarmıştır ama kayda değer bir kalite elde edememiştir[1]. Bu yapıların temel şekilleri Şekil 1.2 'de görülmektedir.

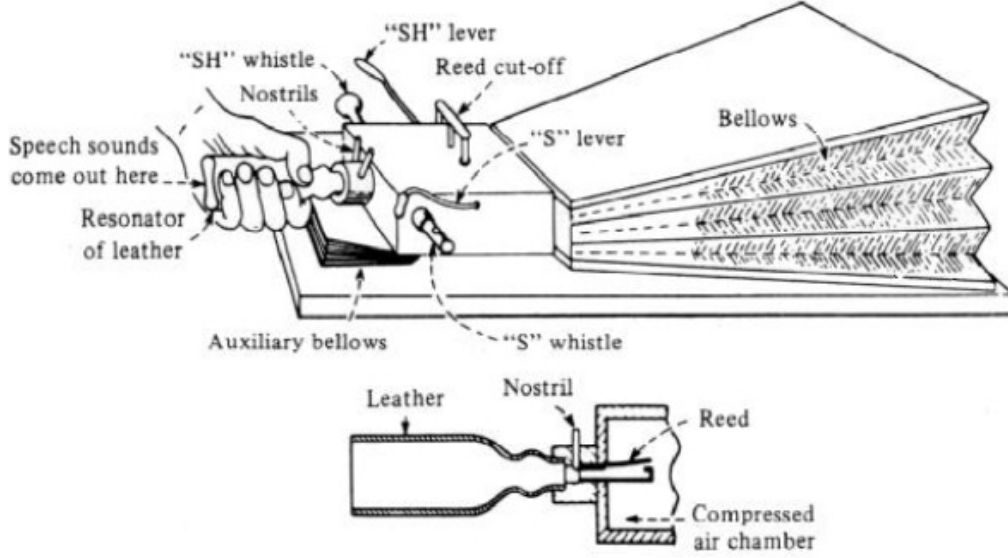


Şekil 1.2 : Kratzenstein'in ses biçimlendirme cihazları.

1791 yılında, Wolfgang von Kempelen "Akustik Mekanik Konuşma Makinesi" adlı bir makine geliştirmiş ve bazı sesleri çıkarmayı kısmen başarmıştır[2]. 1800 lü yıllarda Charles Wheatstone, Kempelen'in cihazını geliştirmiş ve daha iyi sonuçlar elde etmiştir (Şekil 1.3)[3]. Sonraki yıllarda da mekanik tarzda sistemler geliştirilmiş ama kayda değer başarılar elde edilememiştir.

1838 yılında Willis bazı sesli harfler ile ses yolunun geometrisi arasında bir bağlantı kurmuş ve ses yolu yapısına benzeyen borular ile bazı sesli harfleri seslendirebilmiştir. Willis ayrıca seslendirilen seslinin kalitesinin borunun çapına değil sadece uzunluğuna bağlı olduğunu bulmuştur.

Elektrikli tarzda konuşma sentezleyicilerde de çeşitli çalışmalar yapılmış fakat ilk ses sentezleyici olarak Homer Dudley tarafından 1939 yılında sunulan VODER (Voice Operating Demonstrator) kabul edilir[4, 6].



Şekil 1.3 : Wheatstone'un geliştirdiği Kempelen ses cihazı.

İlk Mafsallı(articulatory) Sentezleyici 1958 yılında George Rosen tarafından Massachusetts Institute of Technology 'de yapılmıştır. DAVO(Dynamic Analog of the Vocal Tract) adı verilen tasarım elle kontrol edilen bant kaydı tekniği ile çalışmaktaydı. Noriko Umeda ve arkadaşları tarafından 1968 yılında İngilizce için ilk tam metin seslendirme sistemi Japonya'da tasarlandı. Bu sistem sözdizimsel analiz birimi içeren mafsallı (articulatory) yapıya sahip bir çalışmadır.

İlk Biçimlendirici (Formant) Sentezleyici PAT(Parametric Artificial Talker), Walter Lawrence tarafından 1953 yılında tanıtıldı. 1979 yılında ise Allen, Hunnicutt ve Klatt tarafından biçimlendirici tabanlı MITalk geliştirildi[5].

1970 – 1980 yılları arasında metin seslendirme amaçlı ticari birçok yazılım ve donanım tasarlandı. Daha sonraları DecTalk, Mbrola Whistler gibi farklı diller için uygulamalar geliştirildi.

1.3 Önemli Metinden Ses Üretme Sistemleri

Bu kısımda önemli ve iyi bilinen metinden ses üretme sistemlerinden bazıları kısaca tanıtılacaktır.

1.3.1 MITalk

MITalk Allen, Hunnicutt ve Klatt tarafından 1979 yılında MIT laboratuvarlarında geliştirilmiş biçimlendirici (Formant) temelli bir sentezleyicidir. Bu çalışma, günümüzde kullanılan biçimlendirici temelli birçok çalışma için temel teşkil etmiştir[7].

1.3.2 Infovox

Infovox en tanınmış çok dil destekli ticari metin seslendirme uygulamasıdır. İlk sürümü 1982 yılında İsveç Royal Institute of Technology 'de geliştirilmiştir. Bu sürüm basamaklı biçimlendirici sentezleme yöntemini kullanmaktaydı. Yeni sürümlerinde ise birleştirmeli yöntem kullanılmaktadır. Son sürümlerinde Türkçe de dahil birçok dil desteği sunulmaktadır. Yeni bir ses eklenmesi için şirket kalitesine göre 2-3 hafta ile 3 aya kadar süre talep etmektedir[8].

1.3.3 AT&T Labs Natural Voices & Bell Labs TTS

Bell Laboratuvarlarında geliştirilen bu sistem birim seçme yöntemi uygulayan birleştirmeli bir sentezleme altyapısına sahiptir. Türkçe desteği henüz bulunmayan bu ticari uygulama desteklediği dillerde iyi sayılabilecek sonuçlar vermektedir[9].

1.3.4 ETI Eloquence / Speechworks / Nuance

Eloquence Eloquent Teknoloji (ABD) tarafından geliştirilmiş bağlamalı çok dil desteği sunan bir sistemdir. 1988-2001 yılları arasında çoklu dil desteği sunan ses sentezleme ürünleri üzerine çalışan şirket 2001 yılında Speechworks firması ile birleşmiş sonrasında 2003 yılında Nuance Communications tarafından satın alınmıştır. Şirket seslendirme uygulaması olarak şu anda Realspeak adı altında satılan bir ürün sunmaktadır[10].

1.3.5 MBROLA

MBROLA projesi Belçika Faculte Polytechnique de Mons TCTS Laboratuvarlarında geliştirilmiştir. Projenin amacı çoklu dil desteği veren, ticari olmayan ve araştırma odaklı bir metin seslendirme uygulaması tasarlamaktır. Projede kullanılan yöntem PSOLA(Pitch Synchronous Overlap Add) benzeri bir yöntem olarak tanımlanmaktadır. MBROLA sistemi ikili fonem birleştirme yöntemi kullanmaktadır, fakat MBROLA tam bir metin seslendirme sistemi değildir. Giriş olarak metin değil, gerekli fonemleri, süre ve frekans bilgilerini istemektedir. MBROLA çalışması içerisinde Türkçe de dahil birçok dil için ikili fonem veri tabanı sunmaktadır[11].

1.3.6 Festival TTS

Festival Ses Sentezleme Sistemi Edinburg Üniversitesindeki Ses Teknolojileri Araştırma Merkezi tarafından 90'ların sonunda geliştirilmiştir. Dilden bağımsız ve birçok platformda çalışabilmesi amaçlanan bu proje halen birçok linux dağıtımına da konulmaktadır. Festival sistemi ikili fonem birleştirme yöntemi kullanmaktadır. Festival eğitim, araştırma ve bireysel kullanım için ücretsiz kaynak kodu açık bir uygulamadır[12].

1.3.7 GVZ TTS

GVZ TTS yazılımını Setsek firması tarafından satılan bir ticari uygulamadır. Yöntem olarak birleştirmeli yöntemi kullanmaktadır[13].

2. SES BİLİMİ VE İNSANDA SES ÜRETİMİ

2.1 Türkçe’de Sesler

Dilin temelini sesler oluşturur. Seslerin yazıya çevrilmesinde kullanılan işaretlere harf denir. Türkçede 29 harf vardır. Bunlar sesli, sessiz olmak üzere ikiye ayrılır. Sesli harfler ağızımızdan hiçbir zorlama olmadan çıkan sesleri karşılayan harflerdir. 4’ü kalın 4’ü ince olmak üzere 8 adettir. Ses organlarının söylenişte aldıkları durumlara göre Çizelge 2.1’de gösterildiği gibi sınıflandırılır.

Dilin durumuna göre: Kalın – İnce

Altçene durumuna göre: Geniş – Dar

Dudakların aldığı duruma göre: Düz - Yuvarlak

Çizelge 2.1 : Türkçede Sesli Harfler

SESLİ HARFLER	GENİŞ		DAR	
	KALIN	İNCE	KALIN	İNCE
DÜZ	a	e	ı	i
YUVARLAK	o	ö	u	ü

Sessiz harfler tek başına söylenemeyen, ancak bir ünlü yardımıyla söylenebilen sesleri karşılayan harflerdir. Sessiz harfler 21 tanedir ve bunlar sürekli-süreksiz, sert-yumuşak olarak sınıflandırılabilir.

Sessiz harfler Çizelge 2.2’de görüldüğü gibi sınıflandırılabilir.

Ses tellerinin titreşimine göre: Sert – Yumuşak

Ses yolunun tıkalı ya da açık oluşuna göre: Sürekli – Süreksiz

Sesin çıktığı yere göre: Dudak – Damak – Diş – Gırtlak sessizleri.

Çizelge 2.2 : Türkçede Sessiz Harfler

SESSİZ HARFLER	SERT		YUMUŞAK	
	SÜREKLİ	SÜREKSİZ	SÜREKLİ	SÜREKSİZ
DUDAK	f	p	m , v	b
DİŞ	s , ş	ç , t	j , l , n , r , z	c , d
DAMAK		k	ğ , y	g
GIRTLAK	h			

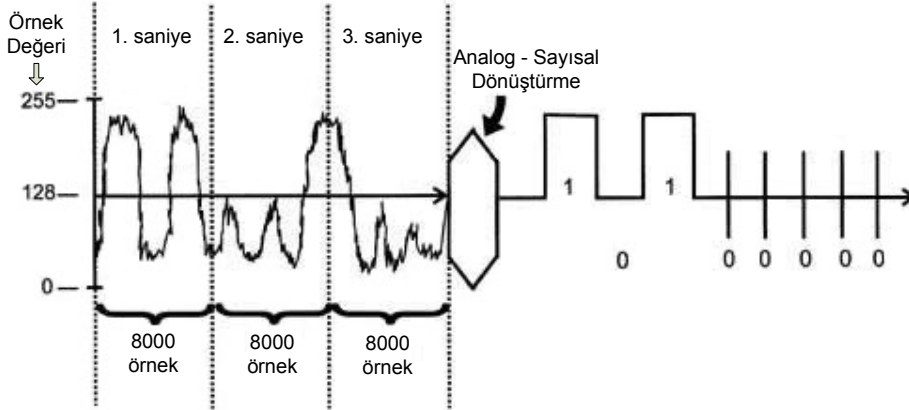
Çizelge 2.3’de Türkçe konuşma seslerinin oluşum şekillerine göre sınıflandırması verilmiştir[14].

Çizelge 2.3 : Oluşum Şekillerine Göre Türkçe Sesler

Ünlüler	Geniş	Dar	Geniş	Dar
Kalın	a	ı	o	u
İnce	e	i	ö	ü
Ünsüzler	Sürtünmeli	Patlamalı	Burunsal	Kaygan
Nefesli	c, j, v, z	b, d, g	m, n	ğ, l, r, y
Nefessiz	ç, f, h, s, ş	t, k, p	-	-

2.2 Sesin Sayısal Kodlanması

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. Şekil 2.1’de ses sinyalinin örneklenmesi özetlenmiştir.

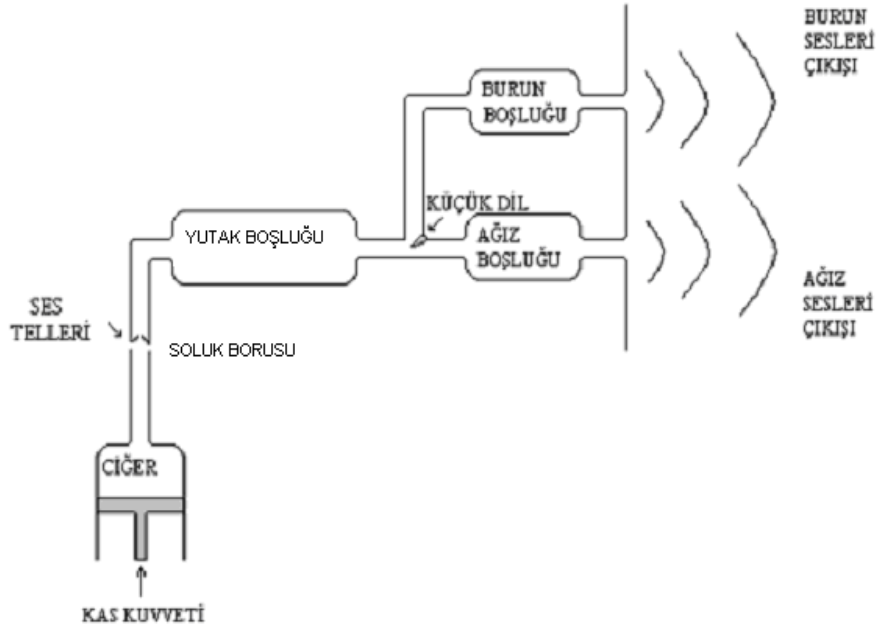


Şekil 2.1 : Ses Sinyalinin Örneklenmesi

Nyquist teoremine göre kayıpsız bir sayısallaştırma için örnekleme oranının frekans yanıtının iki katı ve üstünde bir değer olması gerekmektedir. İnsanın işitsel frekans aralığı 20 Hz – 20 kHz olduğuna göre en fazla 40kHz örnekleme yeterli olmaktadır. Konuşma frekans aralığı ise 200 Hz – 8 kHz aralığında olursa en fazla 16 kHz örnekleme yeterli olacaktır.

2.3 İnsanda Ses Üretimi

Ses dalgası, ses üretim sistemini meydana getiren anatomik yapıların istemli hareketleri sonucunda oluşan, akustik bir basınç dalgasıdır. Bu sistemin ana bölümleri ciğerler, nefes borusu, gırtlak, boğaz, ağız boşluğu ve burun boşluğudur. Ses yolunu oluşturan bu anatomik yapılar, farklı pozisyonlar alarak değişik sesleri oluştururlar. Ses yolunun yapısı Şekil 2.2'de görüldüğü gibi ifade edilebilir. İnsan sesinin dalga şeklini incelediğimizde, dalga şeklinin zamanla değiştiğini görürüz. Konuşma sesleri, kısa süreler boyunca benzer akustik özellikler gösteren ses parçalarına ayrılabilir. Ses sinyallerinin zamana bağlı dalga şekillerine bakılarak, sinyal periyotları, yoğunlukları, süreleri ve her bir ses parçasının sınırları tespit edilebilir. Ancak, ardarda gelen sesler birbirlerini etkilerse bu sınırları belirlemek mümkün olmayabilir. İnsanların ses üretme ve sesi algılama sistemlerindeki organların yapılarından kaynaklanan bazı sınırlamalardan dolayı, ses dalga şeklinde değişime sebep olan yapıların pozisyonlarının kısa zaman aralıklarında değişmediği farzedilebilir[6].

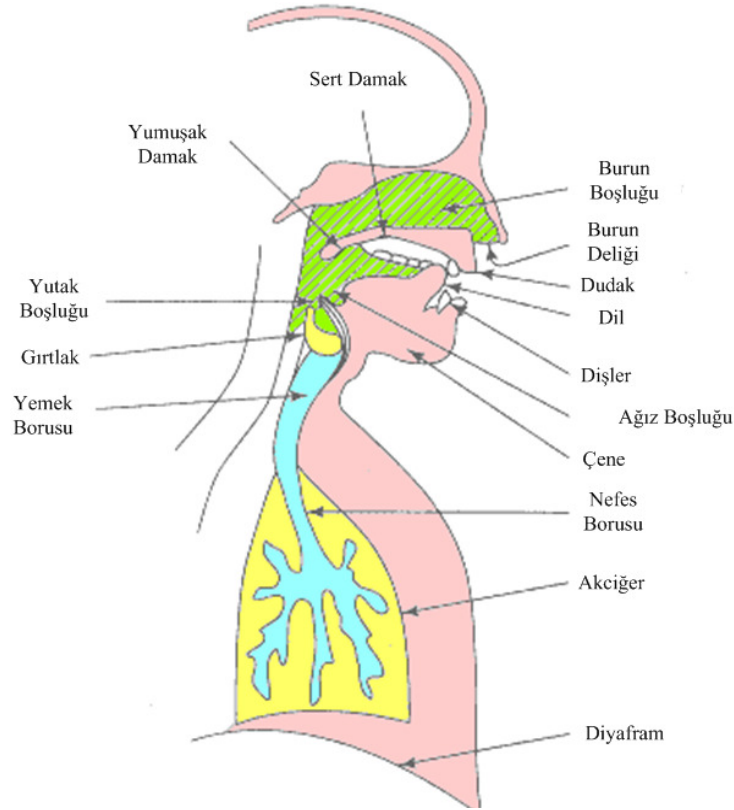


Şekil 2.2 : İnsan Ses Yolunun Yapısı

Konuşma sürekli bir yapı sergilediği ve karmaşık ses sinyalleri içerdiği için yapay olarak sentezlenmesi zordur. Konuşma sinyalleri genellikle ötümlü(voiced) ve ötümsüz(unvoiced) olarak nitelendirilseler de bazen ikisi arasında bir durum da oluşabilmektedir. Ötümlü sesler ses tellerinin titreşiminin temel frekansından (F_0) ve

harmoniklerinden meydana gelmektedir. Ses insan ses yolunda biçimlendirilerek dışarı çıkar. Bu biçimlendirmede temel frekansın yanısıra biçimlendirici(formant) frekansı, genlik ve band genişliği değerleri etkili olmaktadır. Bu biçimlendirme kendini tonlama ve doğallık etkileri üzerinde de göstermektedir.

İnsan sesinin üretilmesine yarayan organlar Şekil 2.3'de görülmektedir[4,14]. Ses üertimi için temel kaynak diyafram ve akciğerlerdir. Hava akımı gırtlak ve ses tellerinden geçerek yutak ve burun boşluğuna ulaşır, sonrasında ağız ve burundan dışarı çıkar. Ses sistemimizdeki en önemli bölge nefes borusunun bitiminde, ses tellerinin arasında bulunan V şeklindeki açıklıktır. Burada ses tellerinde hava akımı modüle edilerek sesli ve nefesli sessiz hafler oluşturulur. Ses tellerinin temel titreşim frekansı erkeklerde 110 Hz, kadınlarda 200Hz ve çocuklarda 300 Hz civarındadır. Gırtlak ve ağız boşluğu arasında yer alan yutağın boyutu gırtlakın yükselip alçalması ile az miktarda değişir. Ağız boşluğunun boyutu ve akustik yapısı damak, dil, dudaklar, yanaklar ve dişlerin hareketi ile değişir. Burun boşluğunun boyutu ise sabit olup buraya giren hava akımı yumuşak damak ile kontrol edilir[4,14].



Şekil 2.3 : İnsan Ses Üretim Organları

2.4 Ses Sentezleme Teknikleri

Sözcük sentezleyiciler giriş olarak, sesbirimine ait parametreler dizisini veya direk olarak metni kullanırlar. Sözcük sentezleyiciler metni, kelimeye ait sözcük giriş yordamlarını kullanarak sözcük birim dizisine çevirir. Büyük sözcük birimleri kullanarak yüksek kalitede sözcük çıkışı elde edilebilir, ancak bunun için yüksek miktarda hafızaya ihtiyaç duyulur. Saklanan sözcük birimleri, konuşma çıkışının birleştirilmesi ile tekrar elde edilir. Konuşma, zamanla değişen doğrusal filtre cevabının modellenmesi (dudaklardan gırtlığa kadar olan kısım) geniş bant gürültüsünden oluşan dalga formu uyarımı ve bir periyodik dalga formu darbelerinin birleşimi ile oluşturulur. Sentezleyici algoritması iki temel yapı altında toplanabilir:

1. Sözcük Saklama Birimi: Konuşma parametreleri saklanıp doğal konuşma elde edilir. Sözcük birimlerinin terimleri belirli bir kurallar dizisi altında toplanır.

2. Birleştirme İşlemleri: Oluşturulan kuralların bir program yardımıyla birleştirilmesi ve bu birimlere ait parametrelerin, zamanla düzgün çıkış elde etmek için ayarlanmasıdır.

Sözcük birimlerinin seçimini, oluşturulması istenen sesin çıkış kalitesi ve gerekli hafıza miktarı kısıtlamaları belirler. Figen Ertaş ve Ömer Eskidere Çizelge 2.4'de gösterilen çizelgeyi sözcük birimlerinin seçimi ile ilgili olarak sunmuşlardır[15].

Sentez sonucunda sözcük çıkışı elde etmek için uygulanan en genel yöntem, insan sesinin kaydedilmesi ve kaydedilen sesin geri okunmasıdır. Kayıt ve okuma işlemi iki farklı şekilde yapılabilir. Analog yapılan kayıtlar pratik ve üretilen ses çok doğaldır. Fakat bu sistem çok büyük hafıza gereksinimi duyar ve pahalıdır. Ayrıca kaydedilen mesajlara seri olarak ulaşılabilmesi bu yöntemin kullanım alanlarını sınırlandırmıştır. Sayısal yöntemler analog yöntemden daha esnek bir yapıya sahiptir. Hafızaya yüklenmiş sözcüklere rasgele ulaşılabilmesi ve sayısal değerler üzerinde değişiklikler yapılması mümkündür. Ancak yine sınırlı hafıza kapasitesi sorunu mevcuttur. Eğer kullanılacak sözcük miktarı çok sınırlı tutulursa, gerçeğine yakın konuşma elde edilebilir. “Sesli Yanıt Sistemleri” (Voice Response Systems) adı verilen sistemlerde bu yöntem uygulanmaktadır. Telefon ile bankacılık, rezervasyon gibi uygulamalar bu türden birer yöntem olup gerçek bir sözcük sentezleyici oluşturulmasından daha basittirler.

Çizelge 2.4 : Ertaş F., Eskiđere Ö. Sözcük Birimlerinin Seçimi Çizelgesi

Sözcük birimleri	Miktar	Tanımlar	Avantajları/Dezavantajları
Kelimeler	300,000 (50,000)	Bir yapının temel birimleri	<u>Avantajı</u> : Yüksek kalite konuşma elde edilir. <u>Dezavantajları</u> : <ul style="list-style-type: none">• Hafıza gereksinimi nedeniyle sınırlıdır.• İzole edilmiş kelimelerin birleştirilmesi gerektiğinden sentetik sözcüğün tabiliği ve ayırt edilebilirliği güçleşecektir.• Kelimelerin sürelerinin iyi şekilde ayarlanması gerekir.
Heceler	20,000 (44,000)	Bir ünlü harf ve yanında ünsüz harflerden oluşur.	<ul style="list-style-type: none">• <u>Dezavantajı</u>: Hecenin sınırları kesin değildir.
Yarım heceler	4500 (2000)	Hecelerin ikiye bölünmesiyle, telaffuzun etkisinin en az olduđu yerde, ünlü harfle kesilerek elde edilir.	<u>Avantajları</u> : <ul style="list-style-type: none">• Sesbirimleri arasındaki geçişler ayarlanıp saklanabilir.• Pürüzsüz ses çıkışı elde edilebilir.
Diphone	1500	Sözcük dalga formlarının sesbirim-alan birimlerine bölünmesi ile elde edilir ve her bir sesbirimi ortasında kesilir.	<u>Avantajları</u> : <ul style="list-style-type: none">• Sesbirimleri arasındaki geçişler ayarlanıp saklanabilir.• Basit pürüzsüzleştirme kuralları kullanılarak pürüzsüz bir sözcük çıkışı elde edilebilir.
Allophone	250	Sesbirimsel deęişikliklerdir.	<u>Avantajları</u> : <ul style="list-style-type: none">• Sesbirimlerinin karşılaştırmalarına baęlı olarak• Algoritma çıkarımlarında komplekslięi azaltır.
Ses birimleri	37	Ses bilimine ait temel birim	<u>Avantajı</u> : Hafıza gereksinimi çok küçüktür. <u>Dezavantajları</u> : <ul style="list-style-type: none">• Telaffuz etkilerinin gösterilmesi için pürüzsüzleştirme amacıyla kompleks kurallara ihtiyaç duyarlar.• Her bir bileşenine ses uyumu oluşturmak için ayarlama yapmak gerekir.

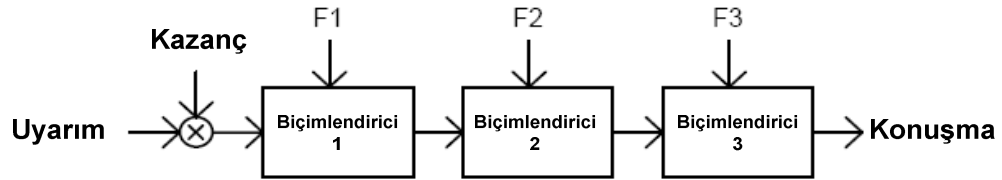
Sentezleyiciler sözcük sentez ve saklama yöntemlerine göre çeşitlilik gösterirler. Bunlardan en önemlileri: Doğal konuşmalardan deęişik birim uzunluklarında yapılan kayıtları kullanan Baęlamalı (Concatenative) Sentezleyiciler, İnsanın ses üretim sistemini modellemeyi amaçlayan Telaffuz ya da Mafsalı Sentezleyiciler (Articulatory Synthesizers) ve ses yolunu kaynak-filtre temelli modellemeyi amaçlayan Formant(Biçimlendirici) Sentezleyiciler (Formant Synthesizers) gibi sınıflara ayrılırlar. Günümüzde en çok kullanılan yöntemler biçimlendirici ve baęlamalı olanlardır. Uzun süre biçimlendirici temelli sentezleme yöntemleri baskın olsa da günümüzde yeni geliştirilen uygulamaların çoęu baęlamalı yöntemleri tercih etmektedirler. Bu yöntemlere aşıęıda deęinilmiştir.

2.4.1 Biçimlendirici (Formant) sentezleyiciler

Muhtemel olarak son on yılda en fazla kullanılan yöntem bu yöntemdir. Bu yöntemle, beyaz gürültünün ya da başka bir sesin, filtre aracılığıyla şekillendirilerek istenilen sesin üretimini amaçlanır. Genel olarak, paralel ve kaskat şeklinde iki temel yapı kullanılır. Fakat en iyi performans, iki yapının çeşitli şekillerde birleştirilmesiyle oluşur[16]. Formant sentezi ayrıca, sonsuz sayıda ses üretimine imkan sağladığı için, diğer yöntemlere göre oldukça esnek bir yapıya sahiptir.

Formant sentezleyiciler, işaretin spektrumundan bulunan formant rezonans ve bant genişliği değerlerini kullanır. Ses yolu filtreleri genellikle 10–23 spektral parametre ile gösterilir. Formant sentezleyiciler LPC sentezleyicilere göre bazı avantajlara sahiptir. Formant sentezleyicide bant genişlikleri parametrelerde kolayca değişiklik yapılarak ayarlanabilir. Sıfırlar, doğrudan doğruya dudak ve dişler gibi filtre özelliği gösterilerek ayarlanabilir. Bununla birlikte sıfır ve kutupların yerleri doğal bir sözcükte otomatik olarak ayarlanır. Ünlü seslerin oluşumu sırasında sesi oluşturan uyarı periyodik uyarı dizisi şeklinde, ünsüz sesler için ise rasgele değişen gürültü şeklinde olmaktadır. Ses yolu kaskat ve paralel rezonatörler kullanılarak modellenir[16].

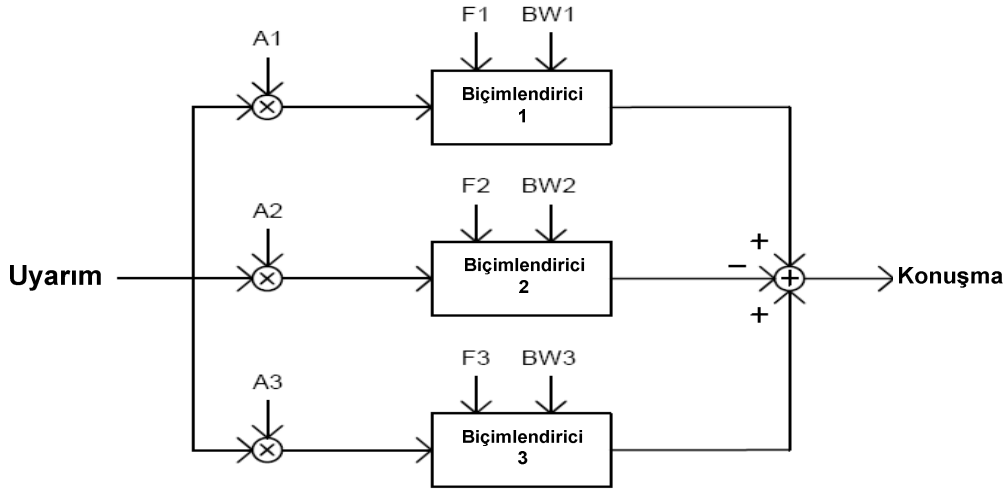
Kaskat rezonatörler genelde ünlü harflerin üretimi için kullanılmaktadır. Kaskat formant sentezi, birbirine seri bir şekilde bağlanmış, bant geçiren yapıda sesi aksettirici birimlerden oluşur. Bir sesi aksettirici birimin çıkışı, diğerinin girişi şeklindedir. Kaskat yapı, kontrol bilgisi olarak sadece formant frekanslarına ihtiyaç duyar. Kaskat formant sentezinde kullanılan temel yapı Şekil 2.4'de gösterildiği gibidir[16].



Şekil 2.4 : Kaskat Biçimlendiriciler

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

Paralel formant sentezinde, ses aksettirici birimler Şekil 2.5’de görüldüğü gibi birlerine paralel olarak bağlanmışlardır. Bazen, genizden gelen sesler için fazladan ses aksettirici birimler de kullanılır. Ses üretimini başlatan sinyal, tüm ses aksettirici birimlere eşzamanlı olarak yollanır ve her bir birimden çıkan sonuç toplanır. Paralel formant sentezinde, hem bant genişliği hem de her bir birimin kazancıyla ilgilenildiği için, daha fazla kontrol bilgisine ihtiyaç duyulur[16].



Şekil 2.5 : Paralel Biçimlendiriciler

Kaskat ve paralel rezonatörler birlikte kullanılarak daha iyi sonuçlar elde edilebilir. Kaskat yaklaşım ile paralel yaklaşımı karşılaştırdığımızda paralel sentezde her bir formant için genlik hesabı yapılması ve formant başına fazladan bir genlik hesabı yapılması gerekir. Bu işlemler ayrıntılı olarak dördüncü bölümde anlatılacaktır. 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[16].

2.4.2 Mafsallı ya da telaffuz (Articulatory) sentezleyiciler

Mafsallı ya da telaffuz sentezleyiciler insanın ses üretim sistemini modellemeyi amaçlamaktadır. Bu açıdan düşünüldüğünde yüksek kalitede yapay ses üretme konusunda ümit vadeden bir yöntemdir. Fakat mafsallı sentezleme yöntemi diğer yöntemlerle kıyaslandığında gerçekleşmesi en zor olan ve en çok işlem gücü gerektiren yöntemdir. Bu nedenle akademik çalışmalar dışında çok fazla üzerinde çalışılmamış ve henüz diğer yöntemlerin başarısına erişememiş bir yöntemdir.

Mafsallı sentezleme genellikle insanın ses sisteminin nefes borusundan ağza kadar olan kısmının modellenmesi şeklinde karşımıza çıkmaktadır. Kural tabanlı sentezleme için 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[17].

2.4.3 Bağlamalı (Concatenative) sentezleyiciler

Doğal sesleri kaydedip, bunları birleştirerek ses sentezlemek, anlaşılır sesler üretmenin en kolay yoludur. Fakat bu kolaylığa karşın, genellikle bir konuşmacı ve bir sesle sınırlanması, diğer yöntemlere de göre daha fazla bellek kapasitesi istemesi gibi olumsuzluklar, seslerin birleştirerek ses üretmenin dezavantajlarındandır.

Kayıtlı sesleri birleştirerek ses üretme yönteminin en önemli aşamalarından biri, doğru birim uzunluklarını bulmaktır. Uzun ya da kısa birimleri seçerken, her iki seçimde de bir şeylerden vazgeçmek gerekecektir. 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. Kısa ses birimleri seçtiğimiz zaman, daha az bellek

ihtiyacı doğmaktadır, fakat örnek toplama ve bunları etiketlendirme daha zor ve karmaşık olmaktadır. Günümüzde kullanılan sistemlerde, yapılan uygulamanın türüne göre kelimeler, heceler, yarım heceler, ses birimleri (fonem) kullanılabilir. Çizelge 2.4’de ses birimlerinin karşılaştırması bulunmaktadır.

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 cümleler karşımıza çıkar. Ayrıca, her dilde, çok sayıda bulunan özel isimlerin varlığı da, metin seslendirme sistemlerinde, ses birimi olarak kelime seçimini uygun kılmamaktadır.

Her dilde farklı olan hece sayısı kelime sayılarına göre oldukça düşüktür. Fakat yine de birçok dil için bu sayı fazladır. Örneğin İngilizce için yaklaşık 10000 hece olduğu kabul edilmektedir. Türkçede ise durum daha olumludur. Türkçe kurallı bir yapıya sahip olduğu için hece sayısı çok daha düşük olmaktadır. Türkçenin bu özelliği hece tabanlı bir sistemin tasarlanması konusunda ümit vermektedir.

Yarı hece kullanımı da Türkçe için hece kullanımı ile aynı şekilde değerlendirilebilir. Hece sayısından çok daha az sayıda yarı hece ile tüm heceler oluşturulabilmektedir. Ayrıca hecede olduğu gibi fonem ya da ikili fonem birleştirmeden çok daha az sayıda birleştirme gerektirmektedir.

Fonem bağlamalı tasarımlarda en çok kullanılan birimdir. Fonem sayısı diğer birimlere göre oldukça düşük olduğu için kural tabanlı sentezlemede büyük esneklik sağlamaktadır. Fakat özellikle patlamalı seslerin çıkarılmasında fonem kullanımı sorunlu olmaktadır. Fonem birimi çoğu zaman diğer birimlerin kullanıldığı sentezleyicilerde de giriş olarak kullanılmaktadır.

3. TÜRKÇE İÇİN METİN SESLENDİRME SİSTEMİ

3.1 Problemler

Türkçe metin seslendirme sistemi tasarımında karşılaşılan problemler sistemdeki işleyiş sırasına göre ele alınırsa önce metin ön işleme aşaması gelir. Sentez aşamasından önce yapılan ön işleme aşamasında girişte alınan ham metin içerisindeki sayı, tarih, kısaltma gibi girdilerin normal metin gibi işleyemeyeceğimizden bunların okunuşunu çıkaran bir birim tasarlayıp bu birimden gelen metin bilgisi sentezleme işlemine geçirilmelidir. Ayrıca ölçülü bir söyleyiş için ön işleme aşamasında bir yapı kurulmalıdır.

3.1.1 Metinden sesçil birimlere dönüştürme

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 ön işleme, doğru söyleyiş için dilbilimsel veri çıkarımı yapılması, doğru tonlama, vurgu ve süre için ölçü analizi yapılması.

3.1.1.1 Metin ön işleme

Metin ön işleme aşaması karmaşık ve dile bağımlı çözümler gerektiren bir aşamadır. Rakamlar, sayılar, kesirler, tarihler, sıra belirten ifadeler, kısaltmalar, özel karakterler sözcüklerle ifade edilir duruma getirilmelidir. Örneğin, 1532 sayısı “binbeşyüzotuziki” şeklinde sentezleyiciye verilmelidir.

Kesirli ifadeler de aynı şekilde okunduğu şekliyle ifade edilmelidir. Örneğin: 4/5 kesir ifadesi “4 bölü 5” ya da “beşte dört” gibi ifade edilerek sentezleme aşamasına geçirilmelidir.

29.10.23, 29.10.1923, 29/10/1923 gibi tarih ifadeleri söylendiği “yirmidokuz ekim bindokuzyüzyirmiüç” şekliyle sentezleyiciye geçirilmelidir.

Sıra belirten ‘.’ karakteri de cümle sonu, kısaltma ya da tarih gibi bir ifade olmadığının anlaşılması ile önündeki sayıya uygun şekilde metne dönüştürülmelidir. “2.”→ ikinci vb.

Kısaltmalar okunuş bakımından çok çeşitli olabileceği için karışıklık olma ihtimali yüksektir. Bu yüzden harf harf söylenmesi daha uygun olacaktır. Örneğin N.A.T.O. kısaltması birçok kişi tarafından kelime şeklinde seslendirilse de harf harf okunacak şekilde dönüştürülmesi karışıklık oluşmasına engel olacaktır. İkinci bir çözüm ise çok sık kullanılan kısaltmaların bir sözlükte tutulması ve dönüştürme işleminin buradan bakılarak yapılmasıdır.

Klavyeden girilebilen özel karakter ise kullanılan dildeki okunuşları ile seslendirilirler. Bu da yeni bir probleme yol açabileceği için uygun düşen karşılıkların yine bir sözlükte tutulması doğru olacaktır.

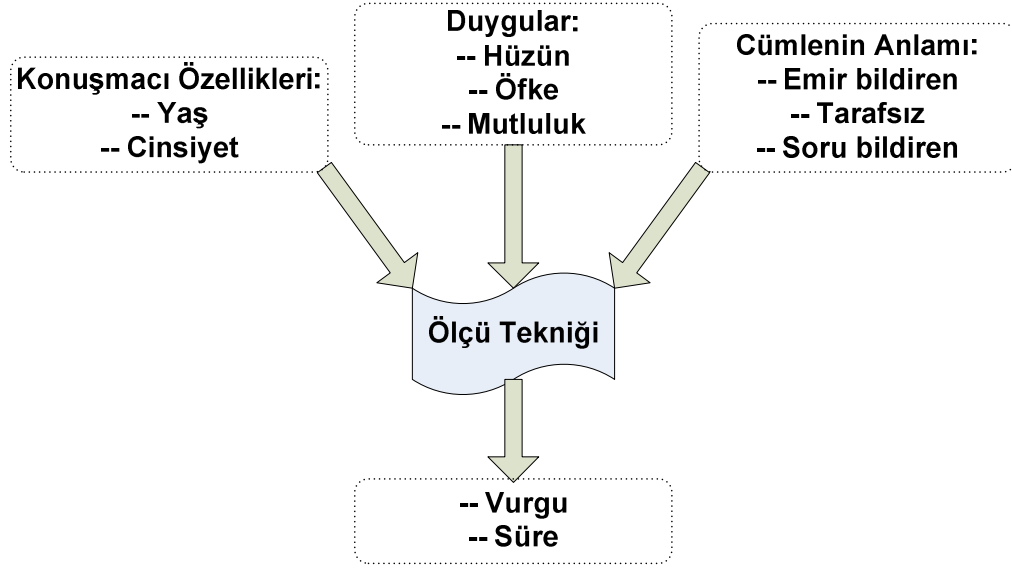
Görüldüğü gibi metin önışlemede her olası durumu doğru bir şekilde çevirmek için karmaşık kurallar tanımlamak gerekebilmektedir.

3.1.1.2 Doğru bir söyleyiş için dilbilimsel veri çıkarımı

Türkçe sesçil bir dil olmasına karşın bazı istisna durumlar sözkonusu olmaktadır. Bu durumların belirlenip bir yöntem geliştirilmesi gerekmektedir. Örneğin babanın kız kardeşi anlamındaki “hala” ile henüz anlamındaki “hala” ’nın söylenişlerindeki farklılık sadece dilbilimsel inceleme ile anlaşılacaktır. Diğer dillerde de aynı durum söz konusu olmaktadır. Sesçil olmayan dillerde ise bu problemler daha fazladır.

3.1.1.3 Ölçü analizi

Söyleyişteki doğru vurgu, tonlama ve zamanlamayı gerçekleştirmek için ölçü analizi yapılmalıdır. Fakat bu analiz halen çeşitli çalışmalarla araştırılan bir konudur. Çünkü ölçü bilgileri Şekil 3.1’de de gösterilen birçok etkene bağlıdır. Fakat bu etkenlerin birçoğu hakkında metinden çıkarım yapmak çok zor ya da imkansızdır. Bu etkiler genelde söyleyiş sırasında ortaya çıkmaktadır.



Şekil 3.1 : Ölçülü Söyleyişi Etkileyen Faktörler

3.1.2 Sentezleme aşaması problemleri

Metinden sesçil birimlere dönüştürme aşamasından sonra elde edilen veriler sentezleme aşamasına verilir. Sentezleme aşamasında da seçilen yönteme göre değişiklik gösteren problemler görülmektedir.

Mafsallı 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 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.

Birleştirmeli ya da bağlamalı yöntemde ise gerekli olan veri tabanını oluşturmak için ses örneklerini toplamak, bunları etiketlemek oldukça uzun zamanlar almaktadır. Ayrıca seçilen ses birimine göre bu veri tabanı oldukça büyük boyutlara ulaşabilmektedir. Ayrıca ses birimlerinin birleştirme noktalarında oluşan bozulmalar anlaşılabilirliği kötü yönde etkilemektedir.

3.2 Türkçe Metin Seslendirme Sistemi

Sesin sentezlenmesi aşamasında kullanılan yöntemle göre metin ve dil analizi yapan birimin çıkışı da değişmektedir. Örneğin sentezleme aşamasında giriş bilgisi fonem listesi şeklinde ise metin analizi biriminin buna uygun çıkışlar verecek şekilde tasarlanması gerekmektedir. Sentezleme aşamasında ise gerekli veri tabanının hazırlanması ve bağlamalı sentezlemeye uygun şekilde birleştirme işleminin gerçekleştirilmesi gerekmektedir.

3.2.1 Ses dosyaları

Ses dosyalarını birçok değişik formatta saklamak mümkündür. Bu projede en çok bilinen formatlardan olan “wav” formatı kullanılmaktadır[18]. Wav dosyasında örnekler ham veri şeklinde sıkıştırılmadan tutulur. Wav dosyası üç veri bölgesi (chunk) içermektedir. RIFF veri bölgesi 12 byte uzunluğundadır ve dosyanın bir “wav” dosyası olduğunun belirtildiği bölgedir. RIFF veri bölgesi alanları Çizelge 3.1’de gösterilmiştir. Sonraki veri bölgesi FORMAT veri bölgesidir. Bu bölgede formata özgü parametreler tanımlanmaktadır ve 24 byte uzunluğundadır. FORMAT veri bölgesi alanları Çizelge 3.2’de gösterilmiştir. Üçüncü veri bölgesi ise DATA veri bölgesidir. Bu alanda gerçek örnekleme verileri tutulur. DATA veri bölgesi alanları Çizelge 3.3’de gösterilmiştir.

Çizelge 3.1 : RIFF Veri Bölgesi

Byte Sırası	Açıklama
0 - 3	"RIFF" (ASCII karakterleri şeklinde)
4 - 7	Little Endian şekilde paketin geri kalanının boyutu.
8 - 11	"WAVE" (ASCII karakterleri şeklinde)

Çizelge 3.2 : FORMAT Veri Bölgesi

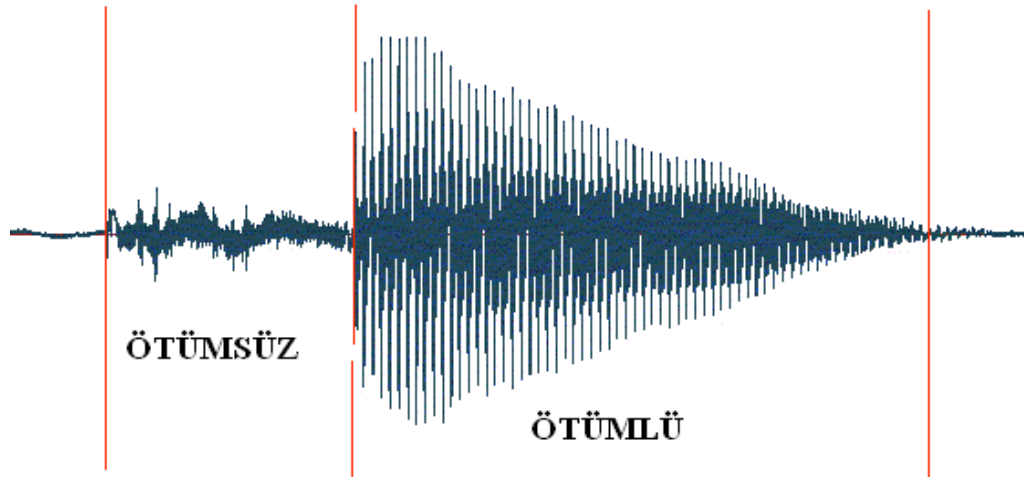
Byte Sırası	Açıklama
0 - 3	"fmt " (ASCII karakterleri şeklinde)
4 - 7	FORMAT bölgesi uzunluğu (Binary, daima 0x10)
8 - 9	Daima 0x01
10 - 11	Kanal sayısı(0x01=Mono, 0x02=Stereo)
12 - 15	Hz olarak örnekleme oranı (binary)
16 - 19	Saniyedeki sekizli miktarı
20 - 21	Örnekteki sekizli miktarı: 1=8 bit Mono,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

Çizelge 3.3 : DATA Veri Bölgesi

Byte Sırası	Açıklama
0 - 3	"data" (ASCII karakterleri şeklinde)
4 - 7	Verinin uzunluğu.
8 - son	Veri (Örnekler)

Wav dosya formatında bir örneğin aldığı değer, o noktada sesin enerji seviyesidir. Bu kodlama şekli Bölüm 2.2’de anlatılmıştır. Buna göre örneğin mutlak değerindeki artış sesin gücündeki artıştır. Sesin gücündeki zamanla değişim wav dosyasında da zamanla değişim şeklinde ifade edilebilmektedir. Tek boyutlu bir dizi şeklinde tutulan örnekleme verisi üzerinde yapılan değişiklik zaman bölgesinde bir değişimi ifade etmektedir.

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



Şekil 3.2 : Ötümsüz ve Ötümlü 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.

3.2.2 Veri tabanı hazırlanması

Veri tabanı hazırlama aşaması bağlamalı sentezleme kullanılan metinden ses üretme sistemlerinin en önemli adımudur. Sistemin çalışabilmesi için gerekli olan ses birimlerinin kayıt edilmiş olması gerekmektedir. Festival ses sentezleme sistemi kullanıldığında seçilen ses birimi ikili fonem olduğu için buna uygun şekilde bütün ikili fonemlerin oluşturulmuş olması gerekmektedir. Hece tabanlı bir sistem için ise

Türkçe için önceden belirlenmiş hece listesinin tamamının kayıt edilmesi gerekmektedir.

Türkçede altı farklı hece türü bulunmaktadır. Bunlar en az bir en fazla dört harften oluşur.

- 1- S → o
- 2- SZ → aç, üç
- 3- SZZ → ilk, üst
- 4- ZS → de, da
- 5- ZSZ → bak, git, beş
- 6- ZSZZ → Türk, dört

Bu altı tip hecenin ilk üç tanesi kelimenin sadece ilk hecesi olabilir. Diğerleri ise kelimenin başında, ortasında veya sonunda bulunabilir. Bu kural dilimize yabancı dillerden gelen sözcükler için sağlanamayabilir. Türkçede hece kuralları belirlidir ve bu kurallı yapı sayesinde toplam hece sayısı diğer dillere göre azdır. Örneğin, Türkçede kelime ya da hece sonuna gelebilecek olan ünsüz çiftleri sadece “-lç, -lk, -lp, -lt, -nç, -nk, -nt, -rç, -rk, -rp, -rs, -rt, -st, -şt” olabilmektedir. Bu durumda Türkçe için olabilecek en çok hece sayısı Çizelge 3.4’de gösterilmiştir. Bu sayı öz Türkçe kuralları dahil edilerek daha da azaltılabilir. Fakat bu durumda yabancı kökenli sözcüklerin söylenebilmesini sağlayacak hecelerde eksikler olacaktır.

Çizelge 3.4 : Türkçe İçin Hece Sayıları

Hece Yapısı	Örnek	Hece Sayısı	Hece Sayısı (Türkçe Kuralları ile)
S	a, e, ı, i, o, ö, u, ü	8	8
SZ	ad, aç, üç, ak, ...	$8 \times 21 = 168$	$8 \times 16 = 128$
SZZ	ilk, üst, ast, ırk, ...	$8 \times 20 \times 20 = 3200$	$8 \times 14 = 112$
ZS	ba, be, de, da, ...	$20 \times 8 = 160$	$19 \times 8 = 152$
ZSZ	bak, git, beş, ...	$20 \times 8 \times 21 = 3360$	$19 \times 8 \times 15 + 4 \times 8 = 2312$
ZSZZ	Türk, dört, ...	$20 \times 8 \times 14 = 2240$	$19 \times 8 \times 14 = 2128$
Toplam		9136	4840

Genel olarak bir hece, tek heceli kelime olabileceği gibi, kelimenin başında, ortasında ya da sonunda olabileceği için bu durumlar için ayrı kayıtlar alınması da gerekebilmektedir.

Veri tabanı hazırlanması sadece hecelerin kayıt edilmesi değil ayrıca her ses dosyası için bir parametre dosyası oluşturulmasını gerektirmektedir. Birleştirme işlemi bu parametre dosyasındaki kesme noktası bildirimlerine göre yapılmaktadır. Parametre dosyasının ilk satırına hecenin dosyadaki başlangıç ve bitiş yeri byte olarak yazılır. İkinci, üçüncü, dördüncü, beşinci ve altıncı satırlarda ise sırasıyla hecenin birinci harfi ön uzatma, birinci harfi son uzatma, ikinci, üçüncü ve dördüncü harflerinin içerisinde peş peşe eklenerek uzatılabilir parçasının başlangıç ve bitiş yerleri byte olarak yazılır.

4. GERÇEKLEME

Hece tabanlı metin seslendirme çalışması iki aşamadan oluşmaktadır. TDK Sesli Sözlük veri tabanı incelenerek birleştirme ve veri hazırlama kuralları çıkarılması, seçilen örneklerin deneme verileri ile oluşturulması için yazılım gerçekleştirilmiştir.

4.1 Birleştirme ve Veri Hazırlama Kurallarının Oluşturulması

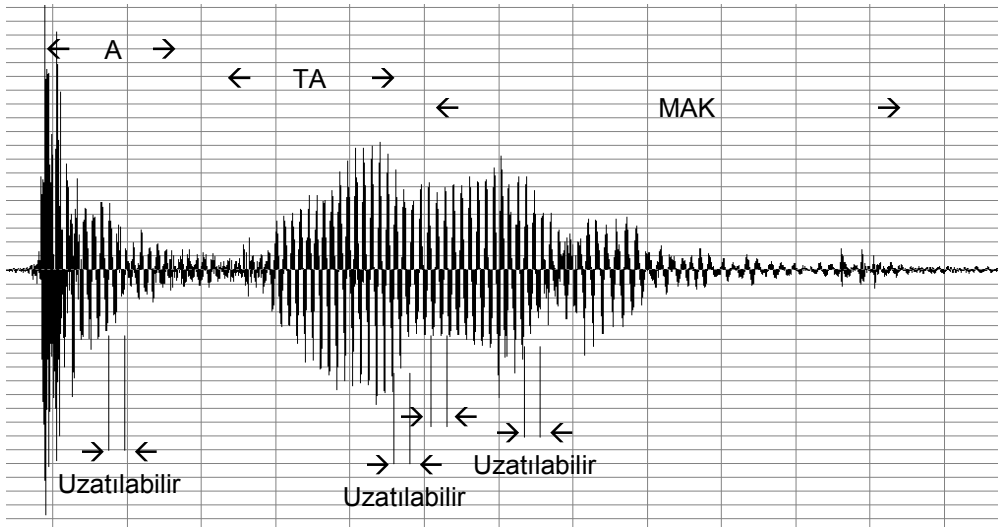
Veri(hece) hazırlama işlemi ve sonrasında uygulanacak birleştirme işlemi için kurallar oluşturulması için kelimelerdeki hecelerin birleşme noktalarında sesler sınıflandırılmaya çalışılmıştır. Yani kelimelerdeki hecelerin birleşme noktalarındaki harflerin sesli-sessiz, nefesli-nefesiz, sürekli-sürekli, ya da oluşum şekillerine göre sürtünmeli-patlamalı-burunsal-kaygan oluşlarına göre nasıl seslendirildiği anlaşılmasına çalışılmıştır. Bu aşamada birçok ses dinlenmiş ve parçalanıp tekrar çalınarak çeşitli kurallar belirlenmiştir. Öncelikle ses verilerinin yani hecelerin mevcut kelimelerden kesilerek alınmasında dikkat edilmesi gereken kuralları açıklamamız gerekir. Bu kurallar aşağıda örneklerle açıklanacaktır.

4.1.1 Hece veri tabanının oluşturulması

- Hece uzunluğunun önemi: Kesilen heceler alındığı kelimedeki farklı kelimelerin oluşturulmasında da kullanılacağı için ortalama uzunlukta heceler seçilmelidir. Özellikle yabancı kökenli kelimelerde uzun söylenen hecelerden alınmaması önemlidir.
- Alınacak olan hece tek başına anlaşılabilir olmalıdır. Ancak kısa heceleri tek başına incelemek zor olduğu için hece kesilerek aynı heceyi içeren başka bir kelimedeki denemeli ve anlaşılır olduğundan emin olunmalıdır. Bu aşamada birleştirme bölgesine işlem yapılmadığı için sese çınlamalar olabilir. Önemli olan hecenin doğru bir şekilde söylenmiş olmasıdır. Örneğin “adalı” sözcüğü dinlendiğinde anlaşılır olmasına rağmen “da” hecesi tek başına dinlendiğinde ya da “adamak” sözcüğündeki “da” yerine eklendiğinde aslında kayıttaki “da” hecesi fazlasıyla hızlı söylenerek çok kısa bir “dı” hecesi gibi

algılanmaktadır. Bu açıdan heceleri mevcut kelimelerden kesme işlemi dikkat gerektirmektedir.

- Kesilen heceler tek kanal(mono), 44100 örnekleme oranı ile PCM(darbe kod modülasyonu) 8 bit olacak şekilde kısa bir dosya süresi seçilerek kayıt edilmelidir.
- Başlangıç, bitiş ve uzatılabilir periyotların işaretlenmesi: Kayıt edilen hecelere ait parametre dosyası için hece başlangıç, bitiş ve hecedeki uzatılabilir seslerin uzatılabilir bir periyodunun başlangıç ve bitiş byte olarak bir dosyaya kayıt edilmelidir. Uzatılabilir periyotları olan sesler, sesli harfler ile burunsal nefesli ve kaygan nefesli sesler olabilmektedir. Şekil 4.1’de başlangıç bitiş ve uzatılabilir bölümler gösterilmiştir.



Şekil 4.1 : İşaretleme işlemi

Hece tabanlı sentezleyici çalışmasında ses dosyaları Türk Dil Kurumu Sesli Sözlük ses dosyaları uygun şekilde kesilerek elde edilmiştir. Sonraki aşamada kayıt yapılarak da hece elde edilmiş ve etiketlenerek denenmiştir. Hecelerin kelimenin başında, sonunda ve ortasında olma durumlarında farklı uzunluklarda olması nedeniyle her hecenin birkaç yerden örnek alınarak denenmesi ve ortalama uzunlukta kayıt altına alınması gereklidir.

Kayıt işlemi tek kanal(mono), 44100 örnekleme oranı ile PCM(darbe kod modülasyonu) 8 bit olacak şekilde yapılmıştır.

4.1.2 Birleřtirme Kuralları

Türkçede Çizelge 3.4’de belirtilen hece yapıları bulunduđuna göre. Hece birleřme noktalarında sadece iki tür birleřme söz konusu olacaktır. Bunlar:

- SZ → Son harfi sesli olan bir hece ile ilk harfi sessiz olan bir hecenin birleřimi.
 - Örn: ata, hece, yazı, v.b.
- ZZ → Son harfi sessiz olan bir hece ile ilk harfi sessiz olan bir hecenin birleřimi.
 - Örn: belge, katkı, güncel, v.b.

Birleřtirme kurallarının çıkarılmasında Goldwave yazılımı kullanılarak TDK veri tabanındaki birçok kelime incelenmiř, heceleri birbirinin yerine yerleřtirilerek uyumlulukları kontrol edilmiřtir. Sonuç olarak ařađıdaki yargılara varılmıřtır:

- Kelimelerde heceler birleřme sınırlarında bazen genlikte bođum oluřturmakta iken bazen oluřturmamakta ya da az oluřturmaktadır. Birleřim yerlerinde bođum oluřturan heceleri ayırt etmek daha kolay olmaktadır(Şekil 4.2 ve Şekil 4.3).
- Bazı heceler kesilerek dođrudan aynı uzunlukta heceye sahip başka kelimelere tařındıđında, bazı kelimelerde çok iyi sonuçlar verirken bazı kelimelerde kötü sonuçlar vermektedir. Bu durum hecelerin birleřim noktalarındaki harflere göre deđiřiklik göstermektedir(Şekil 4.4).

Yapılan deneyler sonucunda bazı kurallar(Çizelge 4.1) oluřturulmaya çalıřılmıřtır. Buna göre hece birleřme noktalarında:

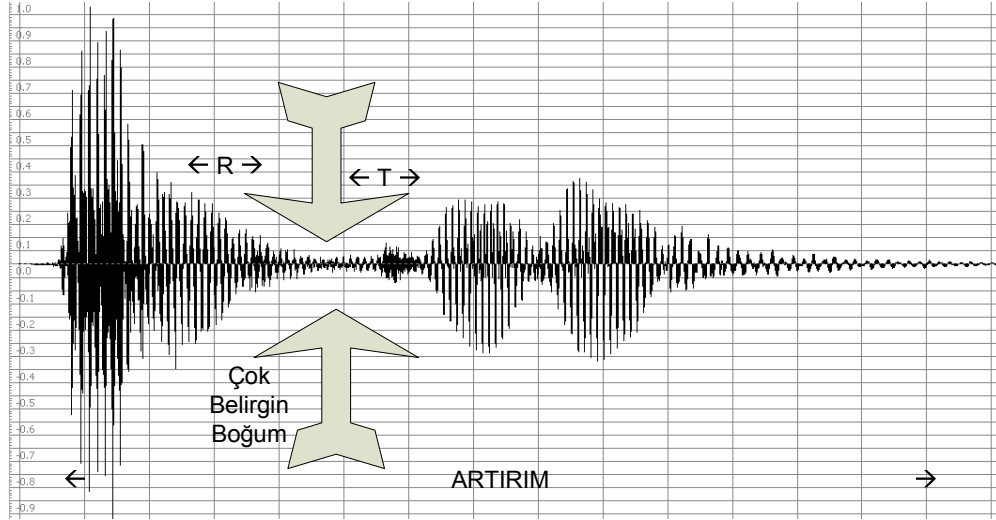
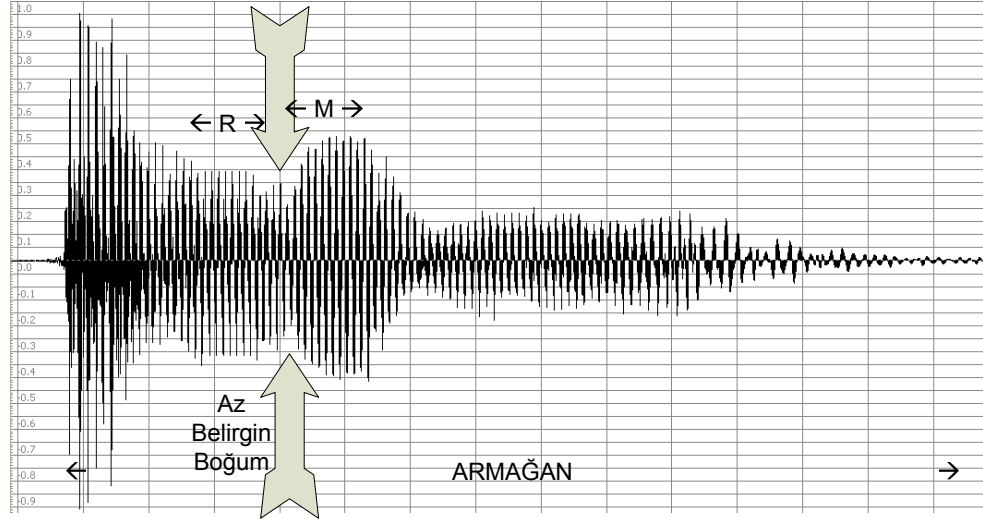
- Sesliden sonra patlamalı sessiz(b, d, g, t, k, p) var ise öndeki sesli harf uzatılır ve bođum oluřturulur. Uzatma 12 periyot kadar yapılmaktadır.(Uzatma ölçüleri çok keskin olmamakla birlikte deneylerde genel olarak aynı büyüklükte uzatmalar iyi sonuçlar vermiřtir)
- Sesliden sonra burunsal(m, n) ya da kaygan(đ, l, r, y) bir sessiz var ise öndeki sesliye uzatma ya da bođumlama iřlemi yapılmadan birleřtirme yapılır.

- Sesliden sonra sürtünmeli+süreksiz (c, ç) bir sessiz gelirse öndeki sesli 4 periyot uzatılır ve boğum oluşturulur. Sürtünmeli+sürekli ise uzatma ya da boğumlama uygulanmaz.
- Sessiz-sessiz (ZZ) durumunda ise, patlamalı ya da nefessiz bir sessizden önce veya sonraki sürekli nefesli sessizde 8 periyot kadar uzatma yapılır ve boğum oluşturulur.

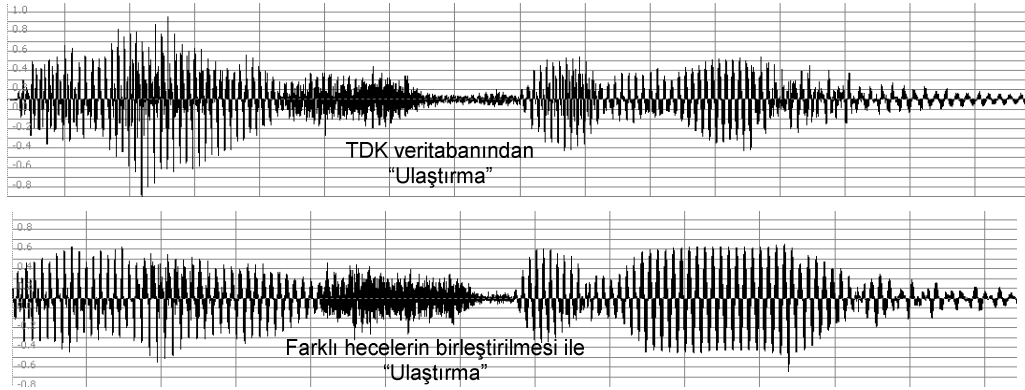
Bu kurallara göre oluşturulmuş örnek dalga formu Şekil 4.3' de görülmektedir.

Çizelge 4.1 : Birleştirme Kuralları

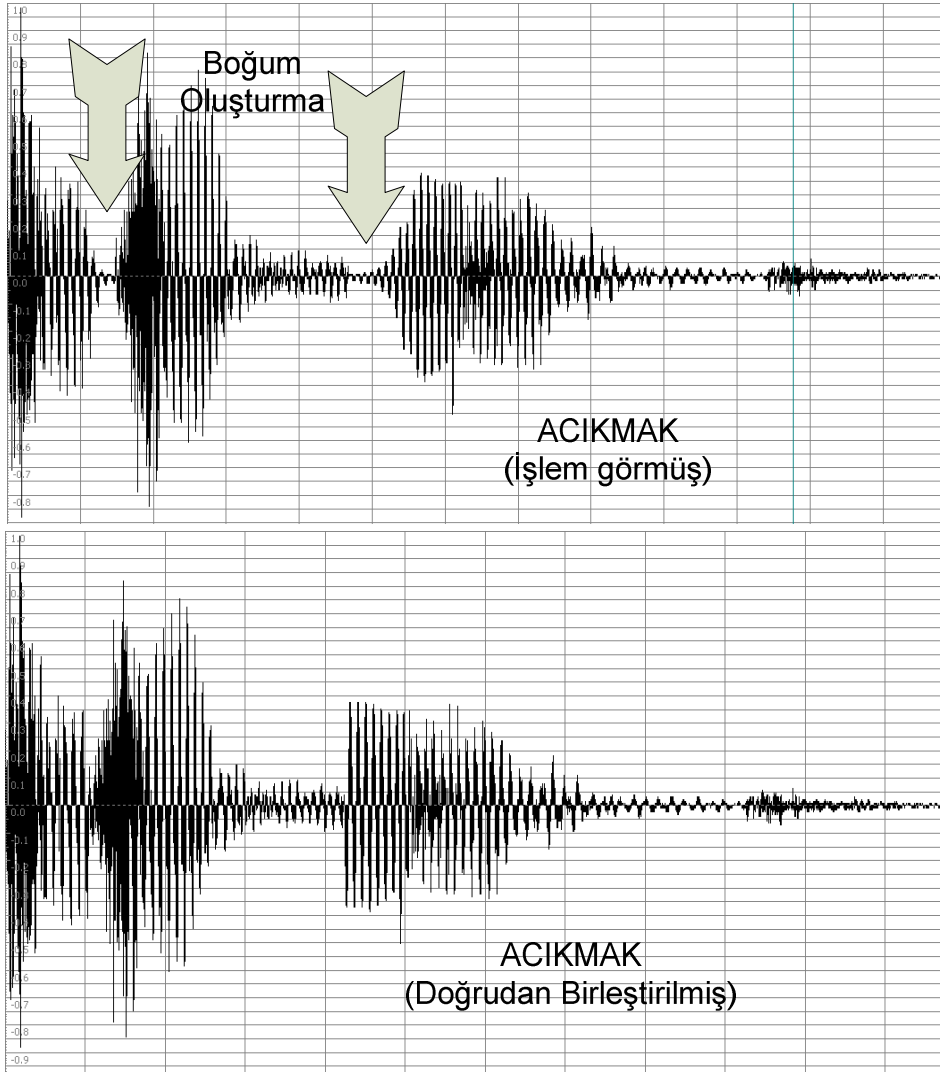
Sesli + Sessiz (SZ)		Sessiz + Sessiz (ZZ)	
S + Patlamalı (b, d, g, t, k, p)	Seslide 12 birim uzatma + boğumlama	Z + Patlamalı ya da nefessiz (b, d, g, t, k, p ç, f, h, s, ş)	Sessiz sürekli ve nefesli (j, v, z, m, n, ğ, l, r, y) ise 8 birim uzatma + boğumlama
S + Burunsal(m, n) ya da Kaygan(ğ, l, r, y)	İşlem yapılmaz	Patlamalı ya da nefessiz (b, d, g, t, k, p ç, f, h, s, ş) + Z	Sessiz sürekli ve nefesli (j, v, z, m, n, ğ, l, r, y) ise 8 birim uzatma + boğumlama
S + Sürtünmeli-Süreksiz (c, ç)	Seslide 4 birim uzatma + boğumlama	-	-
-	-	-	-



Şekil 4.2 : Hece birleşme noktasında boğumlar



Şekil 4.3 : Ulaştırma kelimesinin veri tabanından alınmış gerçek dalga formu ve hecelere oluşturulmuş dalga formu



Şekil 4.4 : Acıkmak kelimesinin kesilmiş hecelerle işlem görmüş ve doğrudan birleştirilmiş dalga formu

4.2 Yazılım Gerçekleme

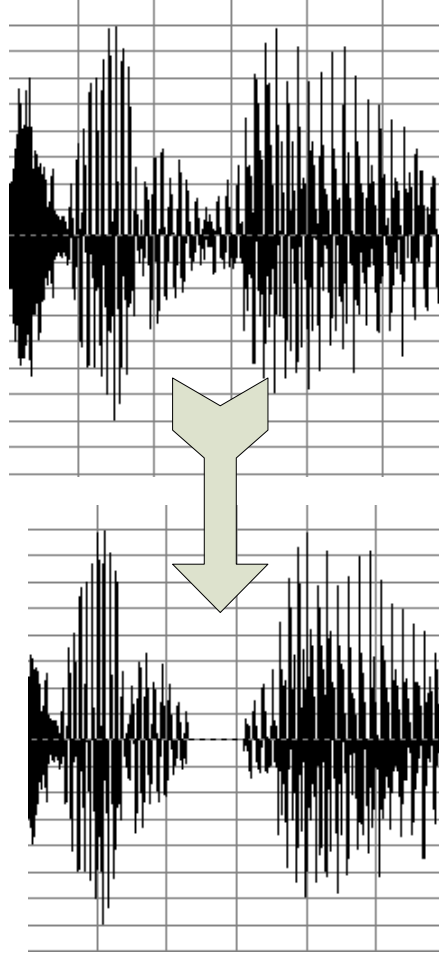
Gerçeklenen sistem kısaca kelimeleri grafik arayüzden alarak bir metin ön işleme yapmaktadır. Bu aşamada Türkçe karakterler Çizelge 4.2’de gösterildiği şekilde UTF-8 karşılıkları kullanılarak işlenmekte, ayrıca girilen nokta, virgül, noktalı virgül gibi bazı özel karakterler yorumlanmaktadır.

Çizelge 4.2 : Türkçeye özgü karakterlerin UTF-8 karşılıkları

ç	0x00E7	Ç	0x00C7
ğ	0x011F	Ğ	0x011E
ı	0x0131	İ	0x0130
ö	0x00F6	Ö	0x00D6
ş	0x015F	Ş	0x015E
ü	0x00FC	Ü	0x00DC

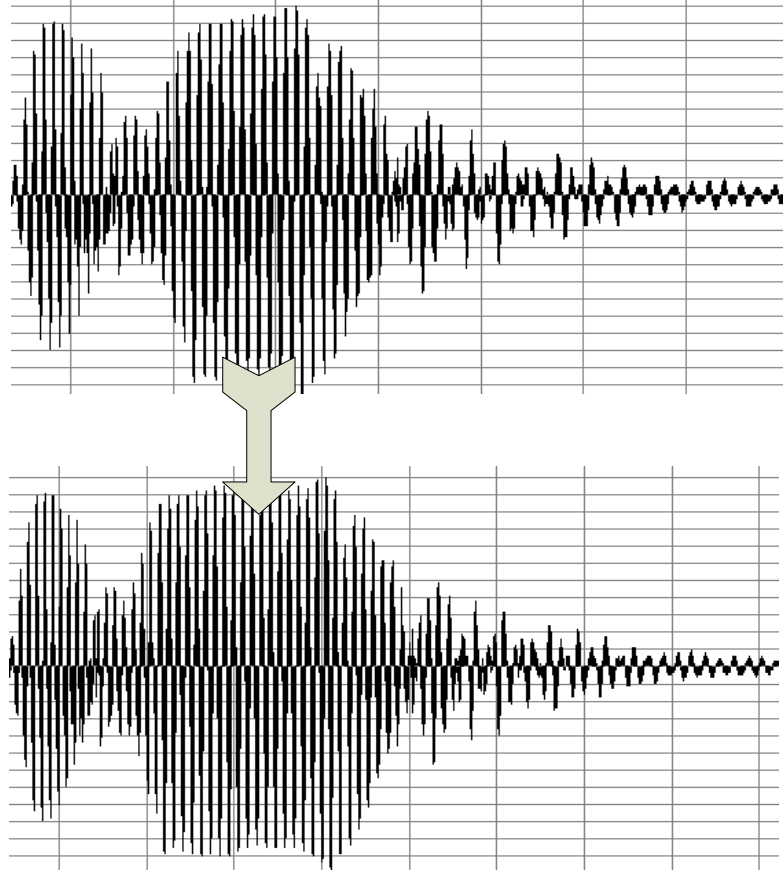
Sonrasında hecelerine ayırma algoritması uygulanmakta ve bulunan heceler kelimedeki yerine göre veri tabanından uygun kayıt ve parametre dosyası alınarak yazılan cümle tamamlanıncaya kadar ekleme işlemi gerçekleştirilmektedir.

Uygun dosyaların alınıp birbiri ardına eklenmesi işlemi okunan parametre dosyasındaki verileri ve birleştirilen heceler birleşme noktalarındaki harf özelliklerini kullanarak süre uzatma, boğum oluşturma ve sessizlik ekleme gibi işlemleri yapar. Bu işlemler vurgu bilgileri tanımlandığı takdirde vurgulu söyleyiş yapmak için de kullanılabilir. Örneğin kelime sınırlarında virgül işaretinden dolayı sessizlik konulması gerekir. Şekil 4.5’de bu işlem gösterilmektedir. Bu işlem nokta ve virgüllerden sonra gerekli olan duraklamanın sağlanması için kullanılmaktadır.



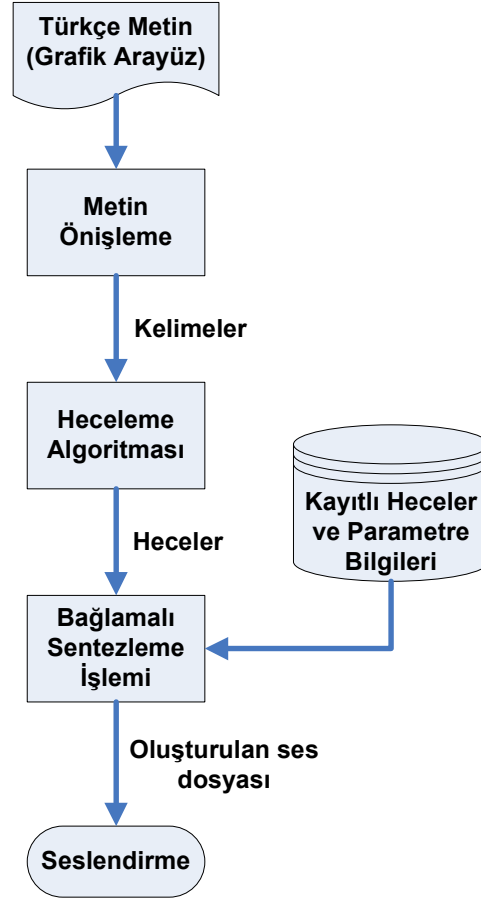
Şekil 4.5 : Birleştirme noktasına sessizlik ekleme

Sessizlik dışında parametre dosyasında bulunan uzatılabilir alanın tekrarlanmasıyla sesli harfler ile sürekli sessizlerde uzatma işlemi yapılabilmektedir. Şekil 4.6'da uzatma işlemi görülmektedir.



Şekil 4.6 : Gerçeklenen Hece Birleştirmeli Sistemde Hece Uzatma İşlemi

Tasarlanan sistemde uzatma işlemi tek başına kullanılmamıştır. Uzatma işlemi boğumlama işleminin ön adımı olarak uygulanmaktadır. Boğumlama işlemi uygulanacak olan hecenin ilgili işaretli kısmında uzatma işlemi boğumlama işlemi ile aynı anda yapılmaktadır. Sonuçta oluşturulan sistemin temel akışı Şekil 4.7’de görülmektedir.



Şekil 4.7 : Gerçeklenen Hece Birleştirmeli Sistem

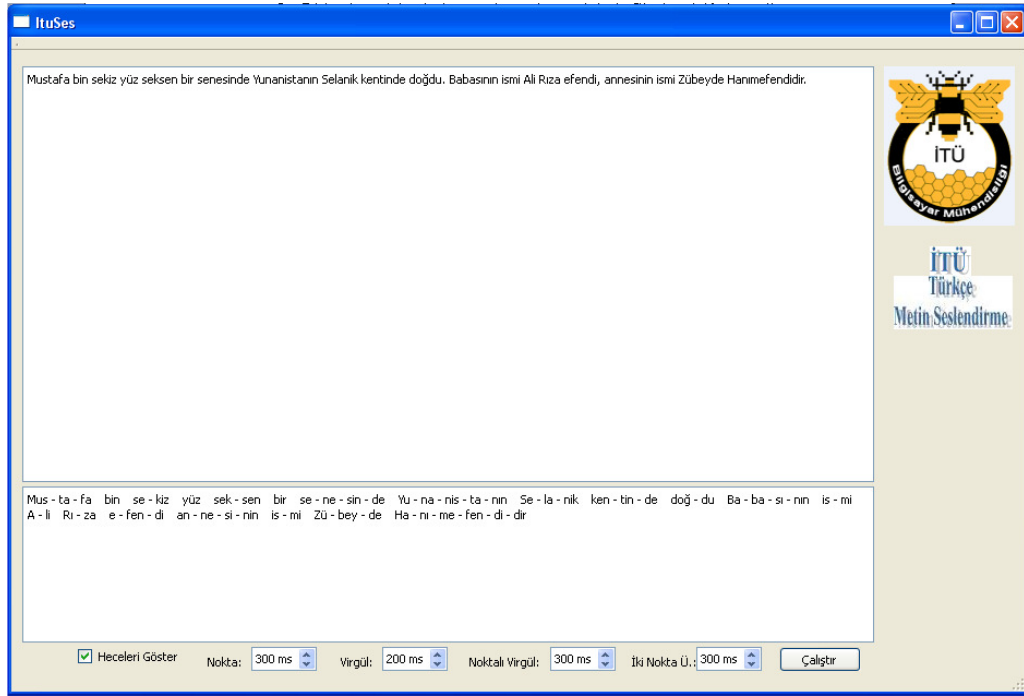
Sentezleme işlemi aşağıdaki adımlardan oluşmaktadır:

1. Bütün metin alınır, çoklu boşluk karakterlerinden ve satır yenilemelerinden arındırılır. Bunların yerine tekil boşluk karakteri yerleştirilir.
2. Elde edilen kelime listesi boşluk karakterlerinden parçalanır.
3. Kelime sonlarında yer alan noktalama işaretlerinin yerleri bellekte tutulur. Sonrasında bu karakterler silinir.
4. Sırayla kelime listesinden kelime alınarak hecelerine ayrılır ve liste yapısında tutulur.
5. Alınan kelime içindeki her hece için bulunduğu yere göre (ilk hece, son hece, ortada), birleşim yeri karar fonksiyonu çağırılır.
6. Bu karar fonksiyonundan dönen değer ve varsa ilgili noktalama işareti bilgisi ile ses dosyası hazırlama işlemi başlatılır.

7. Ses dosyası oluşturma işleminde sese ait etiket dosyasından alınan yer bilgilerine göre ilgili ses dosyası okunur. Alınan veri fonksiyona giren karar tipine ve varsa noktalama işareti bilgilerine göre birleştirilir. Noktalama işaretlerine ait süre bilgileri kullanıcı arayüzünden girilmektedir.
8. Birleştirme işleminde uygulanan karar tipleri, Çizelge 4.1’ de verilen uzatma ve boğumlama işlemleri ile yapılır.

Sistem C programlama dili ve arayüz için QT kütüphanesi kullanılarak gerçekleştirilmiştir. Şekil 4.8’de görülen arayüz kullanılarak giriş yapılabilir.

Sistemin çalışması işletim sisteminden bağımsız olup Ubuntu Linux ve Windows XP işletim sistemlerinde denenmiştir.



Şekil 4.8 : Gerçeklenen Hece Birleştirmeli Sistemin Kullanıcı Arayüzü

5. SONUÇ VE ÖNERİLER

Bu çalışmada, ses birimi olarak heceyi kabul eden bir konuşma sentezleme sistemi geliştirilmiştir. Hece ses çok küçük bir ses birimi olmadığı için insan sesine yakın değerlerde sesler oluşturulabilir. Bağlamalı metin seslendirme sistemlerinde en önemli unsur ses birimlerinin özenli ve doğru alınması ya da kaydedilmesidir. TDK veri tabanında yapılan deneyler sonucunda oluşturulan kurallar çerçevesinde Türk Dil Kurumu veri tabanındaki seslerden kesilerek küçük bir veri tabanı oluşturulmuştur. Oluşturulan bu veri tabanı kullanılarak bağlamalı sentezleme yöntemi kullanılarak Türkçe metin sentezlenmiştir. Sistemin başarısı uygun olarak alınan hecelerin birleştirilmesinde çok iyi sonuçlar verirken, uygun olarak alınamayan hecelerde sadece uyumlu olduğu kelimelerde iyi sonuçlar vermiştir. Oluşturulan veri tabanını kullanarak 50 kelimelik bir paragraf seslendirilmiş, dinleyici görüşleri alınarak Çizelge 5.1'deki sonuçlar ortaya çıkarılmıştır.

Çizelge 5.1 : Dinleyici yorumlarından oluşturulmuş ortalamalar

Ortalama Kelime Anlaşılabilirliği	%75
Kelime İçi Tonlama Başarısı	%65
Cümle Bütününde Tonlama Başarısı	%60
Ses Kalitesi	%60

Geliştirilen sistemde hecelerin birleştirilmesi noktasında meydana gelen tonlama kusurları çalışılmıştır. Bu nedenle cümle bütününde tonlama ve vurgu başarısı çok yüksek değildir. Uygulama geliştirilirken daha çok hece birimleri birleştirme kuralları çıkarılmaya çalışılmıştır. Kayıt kusurlarının giderilmesine yönelik genlik ayarlamaları ya da filtre uygulamaları gerçekleştirilmemiştir. Bu nedenle ses kalitesi düşük ya da ortalama şeklinde yorumlar almaktadır. Ses kalitesinde en yüksek paya sahip unsur doğal olarak hecelerin alındığı veri tabanıdır. Bu veri tabanının uygun ortamlarda sabit ses seviyesi sağlanmaya çalışılarak hazırlanmış olması gerekmektedir.

KAYNAKLAR

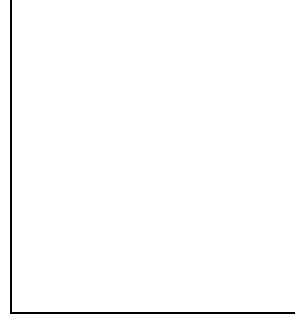
- [1]<<http://www.haskins.yale.edu/featured/heads/SIMULACRA/kratzenstein.html>>, alındığı tarih 20.04.2009.
- [2]<<http://www.haskins.yale.edu/featured/heads/SIMULACRA/kempelen.html>>, alındığı tarih 20.04.2009.
- [3]<<http://www.haskins.yale.edu/featured/heads/SIMULACRA/wheatstone.html>>, alındığı tarih 20.04.2009.
- [4] **Flanagan J. L., Allen J. B., M. A. Hasegawa-Johnson**, 2008. “Speech Analysis Synthesis and Perception”, ch6, ch2.
- [5]<http://www.sarasinstitute.org/Pages/Papers/Fant_Interspeech2005.html>, alındığı tarih 12.04.2009
- [6] **Dutoit T.**, 1997. “An Introduction to Text to Speech Synthesis”, pp 26-32.
- [7] **Dutoit T.**, A Short Introduction to Text-to-Speech Synthesis
<http://tcts.fpms.ac.be/synthesis/introtts_old.html>, alındığı tarih 02.04.2009
- [8]<<http://www.acapela-group.com>>, alındığı tarih 10.04.2009
- [9]<<http://www.research.att.com/~ttsweb/tts/index.php>>, alındığı tarih 01.04.2009
- [10]<<http://www.nuance.com/realspeak>>, alındığı tarih 01.04.2009
- [11] **MBROLA Project Homepage**, <http://tcts.fpms.ac.be/synthesis/mbrola.html>, alındığı tarih 01.04.2009
- [12] **Festival Project Homepage**, <<http://www.cstr.ed.ac.uk/projects/festival/>>, alındığı tarih 01.04.2009
- [13]<<http://www.sestek.com.tr/konusmasentezi.html>>, alındığı tarih 01.04.2009
- [14] **Güner L. Ergenç İ.**, Sesin Doğası ve Oluşumu
http://www.jandarma.tsk.tr/kriminal/turkishinternet/anasayfa/bilarinde_dosyalar/yazilar_dosyalar/bilarinde5.pdf, alındığı tarih 10.04.2009
- [15] **Ertaş F., Eskidere Ö.**, Yazılım Tabanlı Sözcük Sentezleyici DEÜ Fen ve Mühendislik Dergisi Cilt:3 Sayı:1 sh.1-27 Ocak 2001
- [16] **Klatt, D. H.**, 1987: “Review of text-to-speech conversion for English”, MIT,
<http://americanhistory.si.edu/archives/speechsynthesis/dk_737a.htm>, alındığı tarih 01.03.2009

[17] **Lemmetty S.**, “Review of Speech Synthesis Technology” Helsinki University of Technology 1999

[18] **Wave Dosya Formatı**

<<http://ccrma.stanford.edu/courses/422/projects/WaveFormat>>, alındığı tarih 01.04.2009

ÖZGEÇMİŞ



Ad Soyad: Kenan Güldalı
Doğum Yeri ve Tarihi: Erdemli 1980
Adres: Cumhuriyet Mah. Meşecik Sk. No:15/6 Üsküdar
Lisans Üniversite: İstanbul Teknik Üniversitesi
Bilgisayar Mühendisliği