

T.C.  
SAKARYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

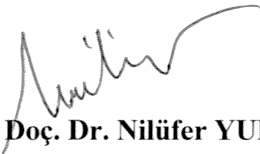
**TÜRKÇE METİNDEN KONUŞMA SENTEZLEME**


**YÜKSEK LİSANS TEZİ**


**Bil.Müh. Yücel BİCİL**

**Enstitü Anabilim Dalı : BİLGİSAYAR - BİLİŞİM MÜH.**

**Bu tez 29/01/2010 tarihinde aşağıdaki jüri tarafından Oybirliği ile kabul edilmiştir.**

  
**Yrd. Doç. Dr. Nilüfer YURTAY**  
**Jüri Başkanı**

  
**Prof. Dr. Mehmed ÖZKAN**  
**Üye**

  
**Yrd. Doç. Dr. Ali GÜLBAĞ**  
**Üye**

## TEŞEKKÜR

Gösterdiği esneklik, bana olan güveni, yol göstericiliği ve verdiği destekle bu eserin ortaya çıkmasına yardımcı olan tez danışmanım Sn. Yrd. Doç. Dr. Nilüfer YURTAY'a ve tez aşamasında hayatını kaybeden hocam Sn. Prof. Dr. Erol EMRE'ye teşekkürü bir borç bilirim. Bu eser, mensubu olmakla gurur duyduğum TÜBİTAK-UEKAE'nin sunduğu teknik imkanlar kullanılarak gerçekleştirilmiştir. Yüksek Lisans çalışmamı bitirmem konusunda destek olan çalışma arkadaşlarım Sn. Salih ERGÜN, Sn. Alper KANAK ve Sn. Mehmet Uğur DOĞAN'a şükranlarımı sunarım. Sentezleme konusundaki tecrübelerini aktaran Sn. Yrd. Doç. Dr. Barış BOZKURT'a ve büyük resmi daima hatırlatan Sn. Prof. Dr. Mehmed ÖZKAN'a ayrıca teşekkür ederim. Her zaman bana inanıp destek olan anneme ve kardeşlerime de teşekkürü bir borç bilirim. Tezin yazılma sürecinde gösterdiği anlayış, sabır ve sevgi için eşim Elife teşekkür ederim. Bu eser yetişmemde büyük rol oynayan babam Mihri BİCİL'e ithaf edilmiştir...

## İÇİNDEKİLER

TEŞEKKÜR.....	ii
İÇİNDEKİLER .....	iii
SİMGELER VE KISALTMALAR LİSTESİ.....	viii
ŞEKİLLER LİSTESİ .....	x
TABLolar LİSTESİ.....	xi
ÖZET.....	xiii
SUMMARY.....	xiv
BÖLÜM 1.	
GİRİŞ.....	1
BÖLÜM 2.	
TARİHÇE.....	
2.1. Mekanik Cihazlar.....	4
2.2. Elektronik Cihazlar.....	5
BÖLÜM 3.	
KONUŞMA SENTEZLEME YÖNTEMLERİ .....	8
3.1. Eklemeli Sentez.....	8
3.1.1. Birim seçme sentezi .....	8
3.1.2. İkili ses sentezleme .....	9
3.1.3. Uygulamaya özgü sentezleme .....	10
3.2. Formant Sentezleme.....	11
3.3. Artikülasyon Bazlı Sentezleme.....	12

3.4. HMM (Hidden Markov Model–Saklı Markov Modeli) Temelli Sentez.....	12
3.5. Mevcut MKS Sistemleri.....	13
3.5.1. MBROLA sistemi .....	13
3.5.2. FESTIVAL MKS sistemi .....	14
3.5.3. MULTEXT sistemi .....	14
3.5.4. GENGLISH sistemi .....	14
3.5.5. HTS sistemi .....	15
3.6. Türkçe MKS için mevcut durum .....	15

#### BÖLÜM 4.

TÜRKÇENİN ANALİZİ .....	16
4.1. Sesbilim .....	17
4.2. Ses organlar.....	18
4.3. Türk Dilinin Yazı Tarihi.....	19
4.4. Türkiye Türkçe'sindeki Seslerin Sınıflandırılması.....	20
4.4.1. Türkçede ünlü sesler.....	20
4.4.2. Türkçedeki ünsüz sesler.....	24
4.4.3. Toklu'ya [34] göre ünsüz seslerin sınıflandırılması.....	25
4.4.3.1. Süreksiz ünsüzler (Kapanma ünsüzleri).....	26
4.4.3.2. Sürekli ünsüzler (Daralma ünsüzleri) .....	26
4.4.3.3. Ötümlü ve ötümsüz ünsüzler.....	26
4.4.4. Özsoy'a [41] göre ünsüz seslerin sınıflandırılması.....	27
4.4.4.1. Oluşma noktasına göre .....	27
4.4.4.2. Oluşma biçimine göre .....	28
4.4.4.3. Gırtlak durumuna göre .....	28
4.4.4.4. Hatipoğlu'na göre ünsüz seslerin sınıflandırılması..	29
4.4.4.5. Ergenç'e göre ünsüz seslerin sınıflandırılması .....	31
4.5. Harf – Ses Ayırımı.....	33
4.5.1. /a/ Sesi.....	33
4.5.2. /e/ Sesi .....	34
4.5.3. /i/ Sesi .....	34
4.5.4. /o/ ve /ö/ Sesleri .....	34

4.5.5. /u/ ve /ü/ Sesleri .....	35
4.5.6. /k/ Sesi .....	35
4.5.7. /g/ Sesi .....	36
4.5.8. /l/ Sesi .....	36
4.5.9. /v/ Sesi .....	37
4.5.10. /r/ Sesi .....	37
4.5.11. /z/ Sesi .....	37
4.5.12. /n/ Sesi .....	37
4.5.13. /ğ/ Olgusu .....	38
4.6. Türkiye Türkçe'sinin Bûrûnsel Özellikleri.....	43
4.6.1. Süre.....	43
4.6.2. Kavşak ve durak.....	44
4.6.3. Ton.....	45
4.6.4. Ezgi .....	45
4.6.5. Vurgu .....	47
4.7. Sonuç .....	50

## BÖLÜM 5.

KONUŞMA SENTEZLEME .....	52
5.1. Genel.....	52
5.2. MKS Sisteminin Aşamaları.....	52
5.2.1. Ses veritabanının hazırlanması.....	53
5.2.2. Doğal dil işleme modülü (Metin ve dil analizi).....	53
5.2.3. Sentezleme modülü (Bûrûn ve konuşma üretimi).....	54
5.3. Bütûncenin Analizi ve Ses Veritabanının Oluşturulması.....	56
5.3.1. Veritabanı tasarımı.....	56
5.3.2. Veritabanının kaydedilmesi .....	60
5.3.3. Veritabanının etiketlenmesi .....	61
5.3.4. Veritabanından, "Sentezleme modelinin" geliştirilmesi.....	63
5.4. Doğal Dil İşleme Modülü.....	63
5.4.1. Metin analizi.....	64
5.4.1.1. Metnin yapısının tespit edilmesi.....	64
5.4.1.2. Metin normalleştirme.....	64

5.4.1.3. Dil yapısının belirlenmesi .....	64
5.4.2. Fonetik analiz .....	65
5.4.2.1. Eş anlamlı sözcüklerin tespiti .....	65
5.4.2.2. Morfolojik analiz .....	65
5.4.2.3. Harften sese dönüştürme .....	65
5.4.3. Bürünsel Analiz .....	66
5.5. Metin Normalleştirme .....	66
5.6. Harften - Sese Dönüştürme Kuralları.....	69
5.6.1. Yöntem.....	70
5.6.2. Deneyler.....	71
5.7. Sentezleme Modülü.....	73
5.7.1. Bürün üretimi.....	74
5.7.2. Konuşma sentezleyici.....	75
5.7.2.1. Kural tabanlı / formant sentezleyici yöntemi.....	75
5.7.2.2. Eklemeli sentezleme yöntemi.....	76
5.7.3. Uygulanan sentezleme yöntemi.....	76
5.7.3.1. Parça seçici.....	78
5.7.3.2. Sentezleyici modülü: kaynaştırıcı.....	78
5.7.3.3. Veritabanı işaret işleme: TDPSOLA referans noktalarının otomatik tespiti.....	79
<b>BÖLÜM 6.</b>	
<b>KONUŞMA SENTEZLEME SİSTEM TASARIMI.....</b>	<b>81</b>
6.1. Kullanıcı Gereksinimleri.....	81
6.2. Sistem Mimarisi.....	82
6.3. Paketler.....	83
6.4. Veri Akışı.....	85
<b>BÖLÜM 7.</b>	
<b>DENEYSEL ÇALIŞMALAR.....</b>	<b>87</b>
<b>BÖLÜM 8.</b>	
<b>SONUÇLAR VE ÖNERİLER.....</b>	<b>90</b>

KAYNAKLAR.....	92
ÖZGEÇMİŞ.....	97

## SİMGELER VE KISALTMALAR LİSTESİ

ASY	: Artikülasyon Sentezleme Sistemi(Articulary Synthesis)
ATR	: Advanced Telecommunication Research Institute
CART	:Sınıflandırma ve Karar Ağaçları (Classification and Regression Trees)
CHATR	: ATR tarafından geliştirilen Sentezleme Sistemi
CSTR	:Konuşma Teknolojileri Araştırma Merkezi (Centre of Spoken Technologies Research)
DDİ	:Doğal Dil İşleme
DECTalk	:Digital Equipment Corporation Sentezleme Sistemi
EST	:Edinburgh Konuşma İşleme Araçları (Edinburgh Speech Tools)
GENGLISH	:Generic ENGLISH
HMM	:Saklı Markov Modeller ( Hidden Markov Models)
HTK	:Saklı Markov Modeller Aracı (Hidden Markov Models Toolkit)
HTS	:Saklı Markov Modeller Sentezleme Aracı ( Hidden MARKov Model To Speech)
IPA	:Uluslararası Fonetik Birliği (International Phonetic Association)
IVR	:Etkileşimli Ses Cevaplama (Interactive Voice Response)
LPC	: Lineer Kestirimci Kodlama (Linear predictive coding)
LTS	:Harften Sese Dönüştürme (Letter To Sound)
MATLAB	: Matematik Laboratuvarı
MBROLA	: Mons Üniversitesi konuşma sentezleme sistemi
MITalk	: Massachusetts Teknoloji Üniversitesi sentezleme sistemi (Massachusetts Institute of Technology)
MKS	:Metinden Konuşma Sentezleme
MKSS	:Metinden Konuşma Sentezleme Sistemi
OCR	:Optik Karakter Tanıma (Optical Character Recognition)



ODTU	:Ortadoęu Teknik Üniversitesi
BILTEN	Bilişim Teknolojileri Enstitüsü
PCM	:Darbe Kod Modülasyonu (Pulse Code Modulation)
PSOLA	:Pitch-Synchronous OverLap and Add
TDPSOLA	:Time Domain Pitch-Synchronous OverLap and Add
TOBI	:Ton ve Durak İndisleri (Tones and Break Indices)
TTS	:Metinden Konuşma Sentezleme (Text To Speech)
TÜBİTAK	:Türkiye Bilimsel ve Teknolojik Araştırmalar Kurumu
UEKAE	:Ulusal Elektronik ve Kriptoloji Araştırma Enstitüsü

## ŞEKİLLER LİSTESİ

Şekil 4.1.	Türkiye Türkçe'sinin ünlü dörtgeni.....	23
Şekil 5.1.	Sadeleştirilmiş MKS Sentezleme süreci .....	53
Şekil 5.2.	MKS Sisteminin genel mimarisi .....	55
Şekil 5.3.	Konuşma işareti üzerinde ikili sesin işaretlenmesi.....	57
Şekil 5.4.	İkili sesler kaynaştırılarak sentezleme yapılması.....	58
Şekil 5.5.	Yarı yansız kayıt odası .....	61
Şekil 5.6.	Wavesurfer kullanılarak yapılan etiketlemeler.....	62
Şekil 5.7.	Doğal dil işleme modülü .....	63
Şekil 5.8.	Sentezleme modülü.....	74
Şekil 5.9.	Konuşma sentezleyici akış şeması.....	77
Şekil 5.10.	TDPSOLA kaynaştırıcı.....	79
Şekil 5.11.	Konuşma işareti TDPSOLA referans noktaları örneği .....	80
Şekil 6.1.	Metinden konuşma sentezleme sistemi akış şeması.....	82
Şekil 6.2.	MKS sistem mimarisi.....	83
Şekil 6.3.	MKSS paket diyagramı.....	85
Şekil 6.4.	MKSS veri akış diyagramı.....	86
Şekil 7.1.	MOS değerlendirme aracı arayüzü.....	91
Şekil 7.2.	TÜBİTAK - UEKAE dinleyici odası görüntüsü .....	91

## TABLolar LİSTESİ

Tablo 4.1.	Türkçedeki ünlü seslerin sınıflandırılması .....	22
Tablo 4.2.	Türkçedeki ünlü sesler ve örnek sözcükler .....	23
Tablo 4.3.	Toklu'ya göre ünsüz seslerin sınıflandırılması.....	27
Tablo 4.4.	Alt ve üst oluşturalara göre ünsüzler .....	28
Tablo 4.5.	Özsoy'a göre ünsüz sınıflandırması .....	29
Tablo 4.6.	Hatipoğlu'na göre ünsüz sınıflandırması .....	31
Tablo 4.7.	/a/ Sesinin durumları .....	33
Tablo 4.8.	/e/ Sesinin durumları .....	34
Tablo 4.9.	/i/ Sesinin durumları .....	34
Tablo 4.10.	/o/ ve /ö/ Seslerinin durumları .....	35
Tablo 4.11.	/u/ ve /ü/ Seslerinin durumları .....	35
Tablo 4.12.	/k/ Sesinin durumları .....	36
Tablo 4.13.	/g/ Sesinin durumları .....	36
Tablo 4.14.	/l/ Sesinin durumları .....	36
Tablo 4.15.	/v/ Sesinin durumları .....	37
Tablo 4.16.	/r/ Sesinin durumları .....	38
Tablo 4.17.	/z/ Sesinin durumları .....	38
Tablo 4.18.	/n/ Sesinin durumları .....	38
Tablo 4.19.	/ğ/ Etkileri .....	40
Tablo 4.20.	Türkçe Ses Kümesinde kullanılan parametreler .....	41
Tablo 4.21.	Türkçe ses kümesi .....	42
Tablo 4.22.	Süre Bilgisi .....	44
Tablo 4.23.	Kavşak Bilgisi .....	44
Tablo 4.24.	Durak Bilgisi.....	45
Tablo 4.25.	Ton Bilgisi .....	45
Tablo 4.26.	Ezgi Türleri .....	47
Tablo 4.27.	Odak Ezgisi .....	47

Tablo 4.28.	Türkçede Üzerine Vurguyu Almayan Ekler .....	49
Tablo 5.1.	Bütünce istatistikleri .....	59
Tablo 5.2.	En fazla geçen üçlü sesler .....	59
Tablo 5.3.	En sık kullanılan üçlü sesler ve logotom listesi .....	60
Tablo 5.4.	Standart olmayan kelimeler ve örnekler .....	68
Tablo 5.5.	Maydanoz kelimesi için elde edilen öznitelik vektörü .....	71
Tablo 5.6.	Türkçe söyleyiş sözlüğü için elde edilen sonuçlar .....	72
Tablo 5.7.	Harf bazlı sonuçlar .....	72
Tablo 5.8.	Konuşma parçacıklarına göre hacim ve kalitesi .....	76
Tablo 6.1.	MKSS kullanıcı gereksinimleri.....	82

## ÖZET

Anahtar kelimeler: Metinden Konuşma Sentezleme, Doğal Dil İşleme, Konuşma Teknolojileri

Bilgisayarların insan hayatında aktif olarak kullanılmasından bu yana gelişen çoklu ortam teknolojileri ve konuşma teknolojileri bilgi ve iletişim çağında önemli rol oynamaktadır. Doğada üretilen veya müziksel enstrümanların ortaya çıkardıkları seslerin yanı sıra insanların ürettikleri sesler, yani insanın konuşması ve konuşulanları anlaması, konuşan kişiyi sesinden tanıması ile ilgili zaman içinde fiziko-akustik modeller geliştirilmiş olup, sinyal işleme teknolojileri de kullanılarak bu modeller bilgisayarlar yardımıyla gerçekleştirilmiştir. Zaman içinde konuşmanın geçtiği dilin iyi bir şekilde modelleyebilmenin de performansı ciddi şekilde etkilediği görülmüş ve doğal dil işleme teknolojileri ile işaret işleme teknolojileri entegre edilerek daha iyi sonuçlara ulaşılmıştır.

Konuşma teknolojileri arasında metinden ses sentezleme konusu oldukça yoğun bir konuşma ve dil işleme çalışması gerektirir. Sentezlemede, sinyal işleme aşamalarından önce, bölütleme ve metin normalleştirme gibi ön işlemlere ihtiyaç duyduğundan, Doğal Dil İşlemenin en yoğun olarak kullanıldığı ses teknolojisi alanı olarak karşımıza çıkmaktadır. Başarılı bir sentezleyicinin geliştirilmesi için ister istemez sentezlenecek metnin dilinin çok iyi analiz edilmesi ve modellenmesi gerekmektedir.

Bu çalışmada Türkçe'nin fonetik ve bürünsel özellikleri incelenmiş ve Türkçe ses kümesi oluşturulmuştur. Türkçe için üçlü seslerin sıklık analizi yapılmış ve en çok kullanılan üçlü sesler belirlenmiştir. Sentezleyicide kullanılacak veritabanı belirlenirken bu üçlü seslerden istifade edilerek hazırlanan veritabanı kullanılmıştır. Yöntem olarak ikili ve üçlü seslerin birleştirilmesine dayanan eklemeli bir sentezleyici geliştirilmiştir. Standart olmayan kelimelerin sentezlenebilmesi için metin normalleştirme birimi geliştirilmiştir. Sentezlenen konuşmaların kalitesinin değerlendirilmesi için öznel bir değerlendirme testi olan Mean Opinion Score (MOS) kullanılmıştır ve bu test çok özel bir altyapı olan yarı yansız. Odalarda ve öznel test odasında icra edilmiştir.

# **TURKISH SPEECH SYNTHESIS**

## **SUMMARY**

Keywords: Text-to-Speech synthesis, natural language processing, speech technologies

Since the actively use of computers and penetration of technologies have risen, multimedia technologies as well as speech technologies play an important role in the information and communication era. In the recent past physico-acoustic models have been developed and combined with signal processing techniques to understand not only the natural non-human voices or musical utterances but also human voice and identify the source of voice or speech. Although the former studies were mostly based on signal processing theory, modeling the spoken language has become a key point that affects the performance of the speech technologies. This has resulted with the efficient combination of natural language processing and signal processing technologies and this strategy has presented better results.

Among the speech technologies, speech synthesis requires a dense work of speech and language processing. In synthesis, before the signal processing phase, segmentation and text normalization are required seriously. This makes synthesis one of the most renowned speech technology that uses NLP dominantly. In order to obtain a well-performing synthesizer, the studied language should be analyzed and modeled appropriately.

In this thesis prosodic and phonetic properties of Turkish are studied and Turkish phoneme set is prepared. The frequency analysis of Turkish triphones is performed and speech database is prepared by taking these triphones into account. The synthesis method is based on an agglutinating method that combines diphones and triphones. In order to synthesize nonstandard words, text normalization is also applied. The quality of the final system is measured in terms of Mean Opinion Score (MOS) in very special subjective test room.

## BÖLÜM 1. GİRİŞ

Konuşma, insanlar arasında iletişimi sağlayan en etkin bir yöntemdir. Sentezleme ise mesajın veya metnin fonetik açılımına göre bilgisayar tarafından “konuşma üretme” işlemine denir [1]. Şu ana kadar geliştirilen sentezleyicilerde anlaşılabilirlik yüksek olmasına rağmen ses kalitesi ve doğallığı halen problem teşkil etmektedir.

Konuşma sentezi teknolojisi, yazılı olan bir metnin sese otomatik olarak çevrilmesi anlamına gelir. Konuşma sentezinin ilk anlamlı örneği 1939 yılında Bell Laboratuvarlarında geliştirilmiş olan Voder sistemidir. ABD'de çocuklara telaffuz öğretmek için Texas Instruments tarafından geliştirilmiş Speak and Spell adlı bir oyuncak ile insanlar bu teknolojiyle tanışmışlardır. Bilgisayarların hafıza ve hızlarının gelişmesi ve teknolojinin de ilerlemesiyle günümüzde oldukça anlaşılır doğala yakın konuşma sentezi kalitesi elde eden sistemler bulunmaktadır.

Konuşma sentezi, ses veritabanında saklanmış ses kayıt parçalarını birbirine ekleyerek yapılabilir. Sistemler kullandıkları parçaların boyutlarına göre farklılık gösterebilirler. Fonemleri veya fonem ikililerini ses parçaları olarak kullanan sistemler lego mantığını kullanarak her türlü kelimeyi az miktarda kayıt kullanarak sentezleme imkanına sahiptirler [2]. Ancak bu tür sentez sistemleri anlaşılabilirlik ve doğallık açısından oldukça zayıftır. Bu bakımdan günümüzde daha uzun ses parçacıklarını kullanan birim seçimli (unit selection) sistemler daha yaygın olarak kullanılmaktadır [3,4].

Eklemeli sistemlere alternatif olarak formant bazlı [5] ve artikülasyon bazlı sentezleme sistemleri de bulunmaktadır. Artikülasyon bazlı Trillium konuşma sentezi

yazılımı gnuspeech adı altında açık kaynak kodlu yazılım projesi olarak süregelmektedir. Bu sistemler için kayıtlı ses parçaları yoktur, sistem ses sinyalini kendisi bir ses üretim modelinin parametrelerini kullanarak üretmektedir. Ancak bu sistemlerin kalitesi de birim seçimli eklemeli sistemlere göre daha düşüktür.

Bu çalışma üçlü sesleri ve ikili seslerin kullanıldığı eklemeli sentez teknolojisi gerçekleştirilmiştir.. Ancak bu sistemlerin de hala uzun metinlerde doğal insan sesine göre dezavantajları bulunmaktadır.

Konuşma sentezi sistemlerinin önemli problemlerinden birisi de metin normalleştirme işlemidir. Metin normalleştirmeden kasıt örnek olarak "269" in "iki altı dokuz" değil de "iki yüz altmış dokuz" olarak okunmasıdır. Bununla ilgili olarak mevcut sistemimiz metin normalleştirme yapmaktadır.

Metinden Konuşma Sentezleme (MKS) sistemi verilen metni konuşmaya çeviren bir sistemdir. Neticede sentezleme sistemi sonucu bir konuşma yani ses dosyası elde edilmektedir. Peki konuşma üretmek veya insanın olmadığı bir ortamda herhangi bir konuşmayı önceden kaydedip te tekrar elde etmek mümkün olduğu halde sentezleme sistemine ihtiyaç nereden çıkmaktadır. MKS sistemi ile diğer sistemler incelenecek olunursa aşağıdaki hususlar göze çaracaktır.

MKS sistemi bütün kelimeleri ve cümleleri sentetik olarak üretirken diğer sistemler önceden kaydedilmiş kelimeleri veya cümleleri birbirine ekleyerek konuşmayı üretir. MKS sistemleri herhangi bir metni kolayca konuşmaya çevirebilir. MKS sistemleri geniş sözlük dağarcıklı bir sistem için kaçınılmazdır. Çünkü geniş dağarcıklı bir sözlükte bütün ihtimallere göre alınacak kayıtlar, hem veri hem de işlem hacmini istenmeyecek seviyelere çıkaracaktır.



Geliştirilecek bir MKS sistemi çok farklı alanlarda kullanılabilir. Bu alanlar aşağıda maddeler halinde sıralanmıştır.

- 1) Telekomünikasyon Servisleri:
  - a) Metinsel bilgilere telefon üzerinden ulaşılmasını sağlar.
  - b) IVR (Interactive Voice Response - Etkileşimli Ses Sistemleri) sistemlerinde karşı tarafa gerekli bilginin okunmasını sağlar.
  - c) Uzaktan metin okuma uygulamalarında kullanılabilir.
- 2) Dil Eğitimi:
  - a) Bilgisayar destekli eğitimlerde yüksek kaliteye sahip MKS sayesinde etkileşimli öğrenme sağlanabilir
- 3) Engelli insanların hayatını kolaylaştırılacak uygulamalar:
  - a) Konuşma engelli kişiler için özel bir klavye ve MKS yardımıyla iletişim kurma imkanı sağlanabilir.
  - b) Görme engelli kişiler için OCR ve MKS sistemleri beraber kullanılarak herhangi bir metni çözümlenmesi sağlanabilir.
- 4) Konuşan kitaplar ve oyuncakların geliştirilmesi
- 5) Anons sistemleri
- 6) İnsan makine etkileşimi

## BÖLÜM 2. TARİHÇE

### 2.1. Mekanik Cihazlar

Elektronik sinyal işleme icat edilmeden çok öncelerde insan sesini elde etmeye çalışan makineler yapmaya çalışanlar vardı. İlk mekanik cihaz örneklerinden “konuşan kafalar” Gerbert of Aurillac (d. 1003 AD), Albertus Magnus (1198–1280), ve Roger Bacon (1214–1294) tarafından yapılmıştır [6].

1779’da Rus Bilim Akademisi’nde çalışmakta olan Danimarkalı bilim adamı Christian Kratzenstein 5 ünlü harf sesini ([a:], [e:], [i:], [o:] and [u:]) üretebilen bir insan ses yolu modeli oluşturdu. Bunun akabinde 1791 yılında Viyana’dan Wolfgang von Kempelen tarafından “Akustik – Mekanik Konuşma Makinesi” tasarımı önerildi. Önerilen bu cihaz ünlü harfler kadar ünsüz harflerin de üretilmesini mümkün kılan dil ve dudak modellemesini de içeriyordu. 1837’de Charles Wheatstone, vonKempelen’in dizaynına dayanan bir “konuşan cihaz” üretti ve 1857’de M. Faber “Euphonia”yı yaptı. Wheatstone’un dizaynı 1923’te Paget tarafından yeniden canlandırıldı [6].

1936 yılında UK telefon şirketi tarafından yapılan konuşan saat uygulaması gerçekleştirilmiştir. Bu sistemde tamlamalar, kelimeler ve kelime parçaları önceden kaydedilip saklanmış olup, daha sonra bunlar bir araya getirilerek arzu edilen konuşma elde edilmiştir.

Aşağı yukarı aynı zamanlarda Bell laboratuvarlarından Homer Dudley, eğitilmiş bir operatörün pedallarını ve mekanik anahtarlarını kullanarak anlaşılabilir bir konuşma

üretebileceği organa benzeyen bir mekanik alet geliştirmiştir. Voder ismi verilen bu alet 1939 yılında Newyork ve Sanfrancisco'daki dünya fuarlarında tanıtılmıştır [7].

Konuşma sinyalinin kaynak filtre modeli ile üretilebileceği gerçeğinden yola çıkılarak analog sistemler de geliştirilmiştir. Gene Homer Dudley tarafından geliştirilen Vocoder bu örneğe girer. 40'lı ve 50'li yıllarda genel konsept konuşma sinyalinin direk kendisinin üretilmesi doğrultusunda olup, fonemlerin üretilmesi ile bu yıllarda çok fazla çalışma yapılmamıştır.

## 2.2. Elektronik Cihazlar

Konuşma sinyallerindeki çalışmalar tanımlanabilir konuşma üretilebilmesi için "Formant Sentezleme" yönteminin geliştirilmesini sağlamıştır. Sinyali doğru bir şekilde gösteren parametrelerin üretilmesi şu an için dahi zor bir süreçtir. İlkel aşamalarda bu parametreler elle ayarlanmış ve otomatik olarak belirlenmesi o zamanlar hedef olarak koyulmuştur. Bugün gelinen noktada parametrelerin doğru bir şekilde ayarlanabilmesi halinde yeterince kaliteli konuşma sentezlenebildiği görülmüştür. Ama tamamen otomatik bir sistemde çok kaliteli bir konuşma elde edilmesi halen zordur.

Bilgisayar teknolojilerindeki gelişmelere paralel olarak işlem gücünün artması kaydedilmiş doğal konuşmalardan birimlerin eklenmesi ile sentezleme üzerinde daha fazla çalışmanın yapılmasını sağlamıştır. Çok kısa bir sürede de ikili seslerin aranılan konuşma birimi olduğu anlaşılmıştır. Ses ikilileri bir önceki fonemin ortasından bir sonraki fonemin ortasına kadar sınırları olan temel yapı birimleridir. Fonem sınırları fonemin ortasına göre çok daha fazla değişken olduğundan fonemin ortası eklem için daha uygundur. 70'li yıllardan beri eklemeli yöntemler üzerinde gerekli disk alanının azaltılabilmesi için çalışmalar yapılmaktadır. Formant sentezleme yöntemi eklemeli yöntemine göre daha az disk alanı istemesine rağmen daha fazla işlem gücüne gereksinim duymaktadır.

1972 de standart unix el kitabı sentezleyici, metin analizi, bürünsel tahmin, fonem üretilmesi süreçleri ile ilgili komutlar içermekte idi. O sıralarda bulunan 16 unix kurulumunun çoğu Murray Hill'deki Bell laboratuarlarında idi.

Teknoloji konuşmayı sıkıştırarak değişik uygulamalar implement edilmesi için geliştirilmekte idi. 70'lerin sonunda Texas Inst. "Speak 'n Spell" oyuncağını piyasaya sürdü. Sesin kalitesi bayağı kötü olsa da o zamanlar için gayet ilgi çekici idi. Konuşma temel olarak LPC ( Lineer Predictive Coding) ile kodlanmış ve genellikle ayrıık kelimeler, harfler ve ekleme ile oluşturulan bazı tamlamalar kullanılmıştır. Belirli çipler üzerine uygulanmış basit TTS (Text to Speech) modülleri ve Apple II gibi ev bilgisayarlarında popülerlik kazanmıştır.

Dennis Klatt'ın MITalk sentezleyicisi bir çok yönden dünyanın ilgisini bu konuya çekti. Daha sonra DECTalk içinde geliştirildi. Robotik ve fakat anlaşılabilir bir konuşma üretmekte idi. Formant sentezleme yöntemini kullanmakta idi [8,9].

80 öncesi konuşma sentezleme para ve zaman harcayabilecek büyük laboratuarlarda çalışılıyor olmasına rağmen, 80'lerin ortalarından itibaren gerekli donanım maliyetlerinin azalmasıyla birlikte bir çok üniversite ve kuruluş ta bu çalışmalara katılmaya başlamıştır. 80'lerin sonunda yazılım tabanlı sentezleyiciler uygulanabilir hale gelmeye başlamıştır.

İşlem gücü ve depolamadaki sınırların ortadan kalkması ile birlikte eklemeli (concatanative) sentezleme için yapılan çalışmalar artmıştır. Japonya da ATR'de çalışan Yoshinoro Sagisaka 80'lerin sonu ve 90'ların başında “nuu-talk” sistemini geliştirmiştir[10]. Bu sistemde daha fazla kaynağa ihtiyaç duyulmakla beraber, aranılan difona (ikili ses) karşılık ses tabanında birden fazla örnek bulunmakta ve otomatik seçim algoritmaları ile en uygun olan difon seçilmektedir. Bu sistem İngilizceye göre çok daha basit fonetik yapısı olan Japonca için geliştirilmiştir. Küçük sestabanlarından kaliteli sentezleme yapmak mümkün olmuştur. 94 yılında

503 cümlelik bir ses tabanından yeni bir parametre dosyası üretmek o zamanki işlemci gücüne göre bir kaç gün almakta ve genellikle sentezleme gerçek zamanda yapılamamakta idi.

Rob Donovan'ın Doktora tezinde İngilizce için tanıttığı “unit selection” sentezleme yöntemi ve 90'ların sonundaki ATR'nin CHATR sistemi [11,12] ile birlikte birim seçme sentezi (unit selection) yoğun çalışılan yöntemlerden biri haline gelmiştir. Sentezlenen çok kaliteli seslerin yanında çok kötü sentezlemelere de bu yöntemde rastlanılmaktadır henüz yüksek ve düşük kaliteli sentezlemeler için %100 güvenli algoritmalar elde edilememiştir.

Konuşma sentezleme konusu diğer konuşma teknolojilerinden bağımsız değildir. Konuşma tanımadaki başarı ve işlem gücündeki artış ve maliyetlerin azalması ile birlikte bir çok kişi işlem gücüne sahip olmuş ve konuşma teknolojileri ile ilgilenmeye başlamıştır.

MBROLA [13] ve Festival [14] gibi sistemler de bir çok insanın kolayca bu teknolojilerin geliştirilmesi sürecine katılmasını sağlamıştır.

## **BÖLÜM 3. KONUŞMA SENTEZLEME YÖNTEMLERİ**

Konuşma sentezleme sistemlerinin en önemli özellikleri doğallık ve anlaşılabilirliktir. Anlaşılabilirlik çıktıların anlaşılma kolaylığı olarak tarif edilirken, doğallık ise çıkan seslerin insan sesine ne kadar yakın olduğu ile ölçülür. İdeal konuşma sentezleyici hem doğal hem de anlaşılabilir olmalıdır. Konuşma sentezleme sistemleri iki özelliği de maksimize etmeye çalışır.

Konuşma sentezleme için başlıca iki teknoloji eklemeli sentez ve formant sentezidir. Her birinin güçlü ve zayıf yönleri vardır. Sentezleme sisteminin kullanım amacı hangi yaklaşımın kullanılacağını belirler.

### **3.1. Eklemeli Sentez**

Eklemeli sentez kaydedilmiş parçaları birbirine eklemeye dayanır. Genellikle, eklemeli sentez en doğal sesli sentezlenmiş konuşmayı üretir. Ancak, konuşmadaki doğal değişimler ve dalga şekillerini bölmelendirmek için kullanılan otomatik teknikler bazen çıktıda duyulabilir küçük teknik problemlere yol açar. Eklemeli sentezin üç ana alt başlığı vardır.

#### **3.1.1. Birim seçme sentezi**

Birim seçme sentezi kaydedilmiş konuşmalardan oluşan geniş bir veritabanı kullanır. Veritabanı oluşturma sürecinde, kaydedilmiş veriler fonemlere, hecelere, kelimelere, cümleciklere ve cümlelere dikkat edilerek etiketlenir. Bu etiketleme işlemi yarı

otomatik olarak yapılabilir. Sphinx [15] ve HTK [16] benzeri konuşma tanıma sistemleri kullanılarak bölütlemeler gerçekleştirilir ve daha sonra bu bölütlemeler elle kontrol edilerek ve ince ayarları yapılarak doğrulanır. Elle düzeltme sürecinde dalga şekli ve frekans spektrumu gibi görsel öğelerden istifade edilir [17]. Etiketlenen veritabanındaki birimlerin indeksi ana frekans (alçalma yükselme açısı), süre, hecedeki yeri ve komşu fonlara dayalı olarak oluşturulur. Sentezleme sürecinde, istenen hedef metin sentezlenirken veritabanından en uygun parçalar (cümle, cümlecik, kelime, hece veya fonemler) seçilmek suretiyle bu parçalar ard arda eklenerek sentezleme işlemi gerçekleştirilir (Parçacık seçme).

Birim seçme sentezi yöntemi ile elde edilen konuşma en doğal sinyaldir çünkü kayıtlı konuşmalara çok az miktarda dijital sinyal işlemleri uygulanır. Bazı sistemler seçilen parçacıkların birbirine eklenmesi sırasında dalga şeklini düzeltmek için az miktarda sinyal işleme yöntemlerinden faydalanırlar. Unutmamak gerekir ki sinyal işleme tekniklerinden ne kadar fazla istifade edilirse, konuşmanın doğallığı o kadar azalacaktır. Veritabanının yeterince geniş tutulduğu sistemlerde elde edilen konuşmalar çoğunlukla gerçek insan sesinden ayırt edilemezler. Ancak, bu kadar geniş bir veritabanının toplanabilmesi çok ciddi emek ve profesyonel spikerlerin kullanılmasını gerektirir. Bazı sistemlerde kayıtlı veriler Gb'larla değişir, bu da saatlerce konuşma demektir [18]. Bu kadar geniş bir veriyi aynı tonlama, aynı genlik ve aynı prozodi ile kaydedebilmek ciddi bir tecrübe gerektirmektedir.

### 3.1.2. İkili ses sentezleme

İkili ses sentezleme bir dilde oluşabilecek tüm ses ikililerini (sesten sese geçişler) içeren minimal konuşma veritabanını kullanır. İkili seslerin sayısı dildeki seslere dayanır. Örneğin; İspanyolcada 800 dolaylarında ses ikilisi bulunurken içerirken Almanca'da bu sayı 2500 civarındadır. Türkçede ise 29 harf olmasına rağmen, seslerin sayısı 44, ikili seslerin sayısı da 2000 civarındadır [19]. İkili ses sentezinde her ses ikilisinin sadece bir örneği konuşma veritabanında yer alır. Yürütme

süresinde cümlenin hedef bürünsel özellikleri bu minimal birimlerin üzerine LPC, PSOLA [20] veya MBROLA [21] gibi sinyal işleme teknikleriyle yerleştirilir. Elde edilen konuşmanın kalitesi genellikle birim seçme sisteminden daha kötüdür ancak formant sentezleyiciye göre daha doğal bir ses elde edilir.

İkili ses sentezleme eklemeli bir sentezleme olmasından dolayı ses geçişlerinde cırtırtılar gibi bazı ses kusurlarını içerebilir. Aynı zamanda LPC gibi sinyal işleme tekniklerinin kullanılmasından dolayı elde edilen sentez robotik olabilmektedir. Ama küçük verilerle gerçekleştirilebilmesi ve bu yöntemle çok fazla sayıda ticari uygulama gerçekleştirilmiş olması halen bu konudaki araştırmaları tetiklemektedir.

### 3.1.3. Uygulamaya özgü sentezleme

Uygulamaya özel sentezlemede, konuşmayı oluşturmak için önceden kaydedilmiş kelimeler ve cümleler sıralanarak birbirine bağlanır. Sistem, metin çıktılarının çeşitliliğinin sınırlı olduğu uygulamalarda kullanılır. Örneğin yol plan anonsları, hava durumu raporları, konuşan saat uygulamaları gibi [22]. Bu teknolojiyi gerçekleştirmek oldukça kolaydır ve konuşan saatler, konuşan hesap makineleri gibi cihazlarda uzun süredir ticari kullanımı mevcuttur. Bu sistemlerin doğallık seviyesi çok yüksektir. Çünkü cümle tiplerinin çeşitliliği sınırlıdır ve sistemin ses çıktıları orijinal kayıtların tonlama ve prozodilerine çok yakındır.

Bu sistemler veri tabanındaki kelime ve deyimlerle sınırlı olduğundan, genel amaçlı sentezlemede kullanılamazlar, sadece önceden programlandıkları kelime ve deyimlerin kombinasyonunu sentezleyebilirler. Ancak doğal konuşma dili içindeki kelime kaynaştırmaları da, çok sayıdaki varyasyonlar hesaba katılmadığı sürece sorun yaratacaktır. Örneğin İngilizcedeki /r/ sesi, “clear” gibi kelimelerde, sadece sonraki kelime sesli harf ile başladığında telaffuz edilir (clear out). Benzer şekilde Fransızcada son sessiz harfler sonraki kelime sesli harf ile başlarsa ulamaya neden olur. Türkçede ise /z/ sesi kelime sonunda ötümsüz, kelime içinde ötümlüdür. Aynı şekilde /r/ sesinin de kelime başında, ortasında ve sonunda söylenişleri farklılık arz



etmektedir [19]. Bu sistemlerde bu gibi durumlara dikkat edilerek veritabanı geliştirilmesi, gerekirse bu istisnaların hepsini kapsayacak şekilde bir veritabanı tasarlanmalıdır.

### 3.2. Formant Sentezleme

Formant sentezleme önceden kaydedilmiş bir ses veritabanını kullanmaz. Bunun yerine sentezlenen konuşma bir akustik model kullanılarak oluşturulur. Temel frekans, ötümlenme (ses tellerinin durumu) gürültü seviyesi gibi parametreler zaman içinde çeşitlendirilerek yapay konuşmanın dalga şekli oluşturulur. Bu metot kural-tabanlı sentezleme olarak da adlandırılır ancak bir çok eklemeli sistem de kural tabanlı bileşenler içerir.

Formant sentezleme teknolojisine dayanan çok sayıda sistem, insan sesiyle karıştırılamayacak kadar yapay ve robotik sesli konuşmalar üretir. Ancak maksimum doğallık her zaman ses sentezleme sisteminin amacı da değildir ve Formant sentezleme sistemlerin eklemeli sistemlere göre avantajları da vardır . Formant sentezlemenin ürettiği konuşmanın anlaşılabilirliği yüksektir, çok yüksek hızlarda bile, ardışıl sistemlerde sık görülen akustik kusurlardan uzaktır. Yüksek hızlı sentezlenmiş konuşma, görme engelliler tarafından, ekran okuyucu kullanarak bilgisayarda hızlı dolaşmada kullanılır. Konuşma örnekleme veri tabanları içermediklerinden, ardışıl sistemlere göre daha küçük boyutludurlar, Bu nedenle bellek ve işlemci gücü sınırlı olan gömülü sistemlerde kullanılabilirler. Formant tabanlı sistemler sentezlenen konuşmanın tüm parametrelerine tam hakim olduğundan, geniş çeşitlilikte vurgu ve tonlama oluşturabilirler. Bunun içinde sadece soru ve ifade çeşitliliği değil, duygu ve ses tonların çeşitliliği de yer almaktadır. Ancak bu parametrelerin doğru hesaplanabilmesi çok karmaşık işlemleri gerektirmektedir.

Formant sentezlemenin, gerçek zamanlı ama yüksek doğrulukta tonlama kontrolüne örnek olarak Texas Instruments firmasının “Speak&Spell” oyuncağı için 1970’lerin

sonunda yaptığı çalışması ve 1980lerin başında Sega'nın arcade makineleri verilebilir [23,24,25].

### **3.3. Artikülasyon Bazlı Sentezleme**

Artikülasyon bazlı sentezleme, sesi insan ses yolunun ve ses yolundaki sesleme süreçlerinin modellerine dayanan hesaplamalı teknikleri ifade eder. Laboratuvar deneyleri için düzenli kullanılan, İlk artikülasyon bazlı sentezleyici Philip Rubin, Tom Baer, ve Paul Mermelstein tarafından Haskins laboratuvarlarında 1970'lerin ortalarında geliştirilmiştir. Bu sentezleyici ASY olarak bilinir ve 1960 ve 1970'lerde Bell laboratuvarlarında Paul Mermelstein, Cecil Coker ve arkadaşları tarafından geliştirilen ses yolu modellerinde dayanır.

Son zamanlara kadar, artikülasyon bazlı sentezleme modelleri ticari ses sentezleme sistemlerinde kullanılmamaktaydı. Bu konudaki önemli bir istisna Calgary üniversitesinin yan şirketi olan Trillum ses araştırma şirketinin geliştirip pazarladığı NeXT tabanlı sistemdir. NeXT'in çeşitli türevlerinin çıkmasından sonra Trillum "GNUspeech" olarak devam etmiştir. 1994'de ilk pazara çıkarılan sistem, tam artikülasyon bazlı metin-konuşma dönüşümünü, Carré'nin ayırıcı alan modeliyle kontrol edilen insan ağız ve burun yollarının dalga boyu kılavuzu ya da iletim hattı analogunu kullanarak sağlamaktadır.

### **3.4. HMM (Hidden Markov Model – Saklı Markov Modeli) Temelli Sentez**

Bu sistemde konuşmanın, frekans spektrumu(ses yolu), temel frekansı(ses kaynağı) ve süresi (bürünsel özellikleri) saklı markov modeliyle eş zamanlı modellenir HMM ile oluşturulan konuşmanın dalga şekilleri maksimum benzerlilik ölçütüne dayanır [26].

### 3.5. Mevcut MKS Sistemleri

TTS sistemleri ilk geliştirilmeye başlanıldığında tek dil düşünülerek geliştirildiğinden ve bu yüzden mevcut bir sistemi diğer dillere çevirmek çok zordu. Yazılım teknolojilerindeki gelişmelerle paralel olarak daha sonra birden fazla dili destekleyecek sistemler geliştirilmeye başlanmıştır. Bu sayede sentezleyicinin geliştirileceği dil için gerekli modeller ve yapı taşları oluşturulduktan sonra mevcut yöntemler çok rahat bir şekilde bu sistemler yardımıyla uygulamaya alınmakta ve geliştiricilerin mevcut dil üzerinde daha iyi modeller oluşturarak daha iyi sonuçlara gitmesi sağlanmaktadır. Bu sistemler aşağıda incelenmiştir:

#### 3.5.1. MBROLA sistemi [27]

Belçika Mons Polytechnic Üniversitesi'nde geliştirilen bir projedir. Projenin amacı akademik araştırmaları desteklemek için ticari olmayan ve çok dilli bir konuşma sentezleyicisini geliştirmek olup PSOLA metoduna benzemektedir. Ticari ve askeri olmayan uygulamalar için MBROLA ücretsiz olarak kullanılabilir. Temelde ikili ses eklemeli (diphone concatenation) bir yöntemdir. Sentezleme için süre ve pitch bilgilerine de ihtiyaç duyduğu için önceden bu bilgilerin önışlemlerle elde edilmesi gerekmektedir. Sentez sonucu üretilen konuşma 16 bit ve 16 KHz kalitesindedir.

Çalışma prensibi olarak ham metin girdisini kabul etmediğinden, fonem , süre ve pitch bilgilerini de gerektiğinden tam bir TTS sistemi olduğu söylenemese de düşük seviyeli bir sentezleyici olarak kullanılması mümkündür.

İngilizce, Brezilya Portekizcesi, Hollandaca, Fransızca, Almanca, Romence, İspanyolca ve Türkçe için difon sestabanı hem erkek hem de bayanlar için mevcuttur.

### **3.5.2. FESTIVAL MKS sistemi [28]**

Edinburgh Üniversitesi, CSTR laboratuvarında Alan Black ve Paul Taylor tarafından geliştirilmeye başlanmış bir projedir. Geliştirme ortamı olarak C++ kullanılmış olup İngilizce, İspanyolca ve Galce için kullanılmaya hazır ses veritabanları ve dil kuralları geliştirilmiştir. Her türlü kullanıma açık bir proje olup, yeni sentezleme metodları üzerinde çalışan, dil modelleri üzerinde araştırma yapan ve alternatif bir TTS sistemi geliştirmek isteyenler düşünülerek dağıtımı yapılmaktadır. Festival LPC, PSOLA ve MBROLA ses veritabanlarını destekleyen bir yapıya sahiptir.

Proje çok dilli sistemleri desteklemek için kaynak kodlarında herhangi bir değişikliği gerektirmeden gerekli script dosyaları ile yeni dil modellerinin ve yeni ses veritabanlarının modüler olarak çalışmasına imkan verir. Dağıtımı yen, linux sürümleri ile beraber yapılmakta olup, windows ortamında çalışma imkanı da sunmaktadır. Skript dili olarak "Schema" kullanılmakta olup her dil modeli için gerekli skriptlerin yazılması gerekmektedir.

### **3.5.3. MULTEXT sistemi [29]**

Provence üniversitesi koordinatörlüğünde geliştirilmektedir. Çeşitli diller için dilbilim araçları, ses veritabanları, dil modelleri geliştirilmesini hedefleyen bir projedir.

### **3.5.4. GENGLISH sistemi [30]**

Bu proje de Mons Polytechnic Üniversitesinde geliştirilen bir projedir. MATLAB için bir "Araç kutusu" (toolbox) olarak geliştirilmiş olup, MKS konusunda çalışacak öğrenci ve araştırmacılara MKS sisteminin bütün aşamalarını "Generic English" adı verilen yirmi kelimelelik bir yapay dil üzerinde göstermeyi hedeflemektedir.

### 3.5.5. HTS sistemi [26]

HMM tabanlı geliştirilmiş olan konuşma sentezleme sistemidir. Nagoya Teknoloji Enstitüsü tarafından dağıtılmakta olup, HTK lisansı ile dağıtılmaktadır. Sistem üzerinde farklı dillerde sentezleme modelleri mevcuttur.

### 3.6. Türkçe MKS için mevcut durum

MBROLA tekniğini kullanan bir sentezleyici Barış Bozkurt [31] tarafından geliştirilmiştir. Ayrıca Boğaziçi Üniversitesi'nde Levent Arslan tarafından bir TTS sistemi geliştirilmektedir. Bu sistem LPC modelleri kullanmakta ve sentezlenecek fonemin sağ ve sol tarafındaki en yakın iki fonemden etkilendiğini öngörerek bir sistem önermektedir. Sentezleme modülü en uygun LPC dosyasından gerekli LPC parametrelerini alır, frekans ve süre bilgilerini de kullanarak konuşmayı sentezler.

Festival projesi kapsamında Sabancı Üniversitesinde Esra Vural [32] ve ODTÜ BİLTEN'de bir grup tarafından Türkçe için çalışmalar yapılmıştır. GENGLISH kapsamında ise Türkçe ile ilgili yapılmış herhangi bir çalışma bulunmamaktadır.

## BÖLÜM 4. TÜRKÇENİN ANALİZİ

Bu bölüm, geliştirilmekte olan Türkçe Metinden Konuşma Sentezleyici için gerekli fonetik altyapı düşünülerek hazırlanmıştır. Kaliteli bir sentez elde edebilmek için dilin fonetik analizinin çok iyi yapılması ve veritabanında kullanılacak bütüncelerin bu analizlere dayandırılması gerekmektedir.. MKS sistemlerinde temel yapı taşları seslerdir ve kelimelerin yazı formlarından (ortografik) söyleyişlerinde kullanılacak seslere açılımının gerçekleşmesi gerekmektedir. Bu bölümdeki bilgiler ışığında Türkçe'nin ses kümesi oluşturulmuş ve harflerden seslere geçebilmek için kurallar belirlenmiştir. Ayrıca sentezlenecek konuşmanın kalitesini artırmak amacıyla kullanılacak bürünsel öğeler de incelenmiştir. Değinilen her konu ne çok fazla yüzeysel bırakılmış, ne de çok fazla derine inmek suretiyle konular arasında kaybolacak derecede incelenmiştir. Halen tartışmaların olduğu konularda uzman görüşleri olduğu gibi aktarılmıştır.

Konunun daha iyi anlaşılması için sesbilim – sesbilgisi ve ses-harf terimleri irdelenmiş, ses oluşumu anlatılmış, şu ana kadar kullandığımız alfabelerdeki seslerden bahsedilmiştir. Türkiye Türkçesindeki mevcut seslerin sınıflandırılmaları ve bürün öğeleri verilmiştir. Bunun yanı sıra Türkiye Türkçesinin ses özelliklerinden bahsedilmiş ve bazı seslerin özel durumları da açıklanmaya çalışılmıştır.

Türkçenin fonetik analizi yapılırken, mümkün olduğunca farklı kaynaklardan yararlanılmaya çalışılmış fakat iki önemli sorunla karşılaşmıştır. Bunlardan birincisi “terim” karmaşasıdır. Bu karmaşa en genel terimlerden, en temel terimlere kadar çok geniş bir yelpazede görülebilmektedir. Bu durumun aşılabilmesi için alternatif terimler mümkün olduğunca verilmeye çalışılmıştır. Diğer önemli sorun ise

Türkiye Türkçe'sinde var olan seslerin sınıflandırılmasıdır. Değişik kaynaklarda Türkiye Türkçe'sinde var olan ünlü ve ünsüz seslerin farklı biçimlerde öbeklenmesi ve sınıflandırmada farklı kriterlerin gözetilmesi karmaşaya yol açmaktadır. Bu durumun aşılabilmesi için farklı kaynaklardan alınan sınıflandırmalar karşılaştırılmıştır ve sentezleyicide kullanılacak sınıflandırma belirlenmeye çalışılmıştır.

#### 4.1. Sesbilim

Sesbilim, konuşurken çıkarılan sesleri inceleyen ve bunların söz konusu dilde doğru çıkışını öğreten bir bilim dalı olup dilbilimin bir koludur [33]. Diğer bir ifade ile sesbilim: Diller üzeri bir yaklaşımla , tüm dillerde insanların kullandıkları konuşma seslerinin neler olduğuna, nasıl oluştuğuna, sınıflandığına, değişik bağlamlarda ve durumlara nasıl değişim gösterdiğine, söylenen herhangi bir sözcenin anlamını aktarmada konuşma seslerinin hangi unsur ve özelliklerinin gerekli olduğuna ilişkin sorulara yanıt arar. Dolayısıyla, yeryüzündeki dillerde kullanılan konuşma seslerinin doğal özelliklerini inceleyen bilim dalıdır [34].

Sesi değişik yönleriyle ele alan iki bilim dalından “Sesbilim” (Fonetik) ve “Görevsel Sesbilim” (Fonoloji) terimlerinin kullanımında bir birlik yoktur. Bazı kaynaklarda Fonoloji için “loji” bilim ekinden dolayı “sesbilim” fonetik için de sesbilgisi kavramları kullanıldığı gibi bunun tam tersi kullanımlara da rastlanmaktadır. Bu konuda bir diğer görüş te şöyledir:

“Dilimizde “sesbilim = Fonetik”, "sesbilgisi = Fonoloji" karşılığı olarak verilmektedir [35]. Kanımızca bilim ve bilgi sözcüklerinin farklı niteliği nedeniyle, özellikle “Fonoloji” karşılığı sesbilgisi terimi doyurucu değildir. “Bilgi” eski deyimiyile “malumat”, “vukuf” bir iş ya da konu üzerinde bilinen şey demektir; “bilim” eski deyimiyile “ilim” ise çevrenin ya da olayların bir kısmını konu edinip deney yollarıyla ve gerçekliğe dayanarak yasalara yükselmeye çalışma anlamına gelir. Bu nedenle “Fonetik”e bilim, “Fonoloji”ye bilgi denemez, her ikisi de birer

bilgi dalıdır. Bunun “Fonetik”in yerini tutan “sesbilim” terimini benimsiyor, “Fonoloji” için doyurucu terim bulunana dek “fonoloji” teriminin kullanılmasını öneriyoruz. Aksi halde verilen karşılıklar ayrı iki bilim dalı arasındaki kesin ayrımı belirtecek güçte olamaz. Bu ayrımı az çok belirleyen “görevsel sesbilim” teriminin, “Fonoloji”nin karşılığı olarak kullanıla gelen sesbilgisi teriminden daha doyurucudur kanısındayız [33].”

#### 4.2. Ses organları

Ses, her şeyden önce havanın titreşmesi sonucunda oluşan ve kulakla duyulan fiziksel bir olaydır. Dil seslerinin oluşmasında öncelikle akciğerlerden gelen havaya gereksinim vardır. Ciğerlerin sıkışması ile yükselen hava akımı dışarı çıkarken gırtlakın sağ ve sol tarafında bulunan ses tellerine çarpar ve tınıya/ salt sese dönüşür. Gırtlakta oluşan gürültüler, konuşma sesi değildir (örneğin dil sesinden farklı olarak düzensiz ve süreksizdirler).

Bu tınların konuşma sesine dönüşmesi için titreşim bölgelerinden biri veya daha fazlasından burun boşluğu, ağız boşluğu ve gırtlak boşluğundan geçmesi gerekir. Bu üç boşluktan biri olan ağız boşluğu, gırtlakta oluşan güzel bir ses için gerekli özellikleri taşımayan salt sesleri konuşma sesine dönüştürmek için gerekli organları bulunduran en önemli titreşim bölgesidir. Ağız boşluğu üst ve alt olmak üzere iki bölümden oluşmaktadır. Ağız boşluğunun üst kısmında önden arkaya doğru üst dudak, üst dişler, üst dişlerin gömülü olduğu dişeti, et damak (ağız boşluğunun çatısının ön kısmı), yumuşak damak (ağız boşluğu çatısının arka kısmı) ve küçük dil (bazı kaynaklarda uvula diye geçer) bulunur.

Ağız boşluğunun alt kısmında sırasıyla altdudak, alt sıra dişleri ve dil bulunmaktadır. Alt ve üst sıra dişleri konuşma seslerinin oluşmasında çok önemli görevler üstlenmektedir. Ancak, seslerin söylenişi sırasında dişler her zaman edilgen organ rolünü üstlenmektedir (yani hareket etmeyen ve ötekilerden az hareket eden



organlardır). Dil güçlü kaslarla örtülü, bir çok biçim alabilen, bu yüzden de seslerin şekillenmesinde etkili olan organdır. Dili değişik bölümlere ayırabiliriz: uç, ön, dil palası, orta, arka ve kök. Dilin uç kısmı ses aygıtının en hareketli, en atletik kısmıdır. Bu nedenle konuşma seslerinin oluşmasında en sık kullanılan eklemleyicidir [36].

### 4.3. Türk Dilinin Yazı Tarihi:

Türk Dili, sırasıyla Göktürk (Orhon), Uygur, Arap ve Latin alfabeleriyle yazılmıştır. Göktürk alfabesinde dördü ünlü 38 harf/işaret vardır. Türkçe'deki 8 ünlü için Göktürk alfabesinde sadece 4 im kullanılmıştır: /a/ ve /e/, /ı/ ve /i/, /o/ ve /u/, /ö/ ve /ü/ aynı harflerle simgelenmiştir. Ünlüler için durumun böyle olmasına karşılık, bazı ünsüzler için ( /b/, /d/, /g/, /k/, /l/, /n/, /r/, /s/, /t/, /y/ ) ikişer harf kullanılmıştır.

Uygur alfabesi sadece 18 harften oluşmuştur. Bu harflerden 3 tanesi ünlü, geri kalan 15 harf ise ünsüzdür. Uygur alfabesi sağdan sola doğru yazılan ve harflerin biçimleri sözcüklerin başında, ortasında ve sonunda değişen bir sistemdir.

Daha sonra kullanılan Arap alfabesi de 33 harfliktir. Klasik 29 harflik Arap alfabesine ek olarak Türkçe'nin ses düzenine daha uygun olan ve İranlıların Arap alfabesine Farsçadaki ek dört sesi simgelemek için 4 harf (/p/, /ç/, /j/, /g/) ekledikleri 33 harflik alfabe kullanmışlardır.

1 Kasım 1928'de kabul edilen ve bugün kullandığımız Latin alfabesinde ise Türkiye Türkçe'sindeki sesler 29 harfle gösterilmektedir. Bu harflerden sekizi ünlüleri, diğerleri de ünsüzleri göstermektedir. Bu yeni alfabe İstanbul ağzını örnek alan sesçil/fonetik bir alfabadır.

#### 4.4. Türkiye Türkçe'sindeki Seslerin Sınıflandırılması

Türkiye Türkçesinde sesler ünlü ve ünsüz olmak üzere iki ana sınıfa ayrılır. Ünsüz sesler çıkarılırken, soluğumuz ses yolunda dil, diş ve dudak gibi bir engelle karşılaşılır. Engellerle karşılaşma biçimlerine göre de alt sınıflandırılmaya gidilir. Ünlü sesler çıkarılırken ise soluğumuz herhangi bir engelle karşılaşmaz. Ses yolunun açıklık derecesi ünlülerin çeşitlerine göre değişir. Bu değişiklik çenelerin az veya çok açılması, dilin kabarıp alçalması ve dudakların yuvarlaklaşıp düzleşmesi ile temin edilir [37].

Türkçe, ünlü açısından çok zengin ve ünsüz açısından da oldukça çeşitli bir dildir. Türk alfabesinde 8 ünlü, 21 ünsüz olmak üzere 29 harf bulunur. Fonetik açıdan ise 15 ünlü, 29 ünsüz olmak üzere 44 fonem mevcuttur ki bu, Türkçe'nin popüler olan diğer dillere göre çok zengin bir fonetik yelpazeye sahip olduğunun göstergesidir [40].

##### 4.4.1. Türkçede ünlü sesler

Türkiye Türkçesinde ünlü sesler vokaller [37], sesliler [38] veya ünlüler (vowel) [33,39] olmak üzere değişik terimlerle ifade edilmektedir. Her ne kadar ünlüleri isimlendirirken terminolojide uzmanlar arasında bir uzlaşma söz konusu olmasa da sınıflandırmada bir karmaşa bulunmamaktadır.

Ünlü sesler, oluşumları sırasında herhangi bir engelle karşılaşmadan yalnızca ses tellerinin titreşimiyle üretilen ve ağzın değişik durumlarından tını ve renklerini alan seslerdir[34]. Ünlüler özgür ve gürültüsüz seslerdir. Bir diğer tanımla akciğerlerden gelen ve soluğun hiçbir sürtünmeye ve engellemeye uğramadan dilin üstünden geçerken çıkardığı seslerdir. Bu sesler çıkarılırken konuşma organlarının herhangi bir yerinde kapanma ya da daralma olmadığı için gürültüsüzdürler [39].

Ünlülerin çıkarılışında en önemli görevi üstlenen organlar dil, çene ve dudaklardır. Bu nedenle ünlülerin sınıflandırılması bu organlara göre yapılır.

Çene Açısının Durumuna göre:

Geniş Ünlüler : /a/, /e/, /o/, /ö/

Dar Ünlüler : /ı/, /i/, /u/, /ü/

Dudakların Biçimine Göre:

Düz Ünlüler : /ı/, /i/, /a/, /e/

Yuvarlak Ünlüler : /o/, /ö/, /u/, /ü/

Dilin Devinimine Göre:

Arkadil Ünlüleri : /ı/, /a/, /o/, /u/

Öndil Ünlüleri" : /i/, /e/, /ö/, /ü/

Öndil ünlüleri de kendi aralarında ikiye ayrılır:

Yuvarlak Öndil Ünlüleri : /ö/, /ü/

Düz Öndil Ünlüleri : /e/, /i/

Yukarıdaki bilgiler ışığında oluşturulan Türkçe'nin ünlü sınıflandırılması Tablo 4.1'de verilmiştir.

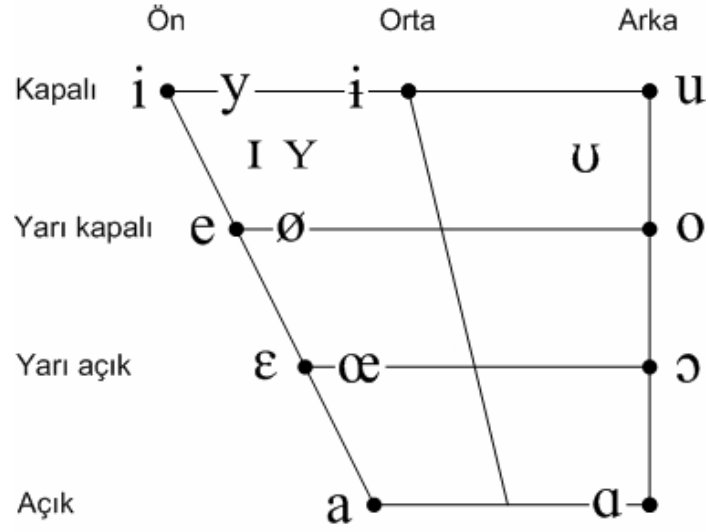
Tablo 4.1. Türkçedeki ünlü seslerin sınıflandırılması

		Dilin Durumuna Göre			
		Öndil Ünlüleri		Arkadil Ünlüleri	
Çene Açısının Durumuna Göre		Dar	Geniş	Dar	Geniş
Dudakların Durumuna Göre	Düz	/i/	/e/	/ɪ/	/a/
	Yuvarlak	/ü/	/ö/	/u/	/o/

Her ne kadar yukarıdaki sınıflandırma harf bazında doğru olsa da söyleyiş esnasında ünlüleri temsil eden harflerin birden fazla seslemleri bulunmaktadır. Örneğin,

1. Arkadil ünlüleri arasında yer alan /ɪ/ ünlüsü çıkış açısından ortadil ünlüsü olarak ta görülebilmektedir [33] ve tek bir durumu vardır [39].
2. /a/ sesinin öndamaksıl (predorsal) ve artdamaksıl (postdorsal) olmak üzere iki farklı durumu vardır [39].
3. /ɪ/ ve /a/ dışında kalan bütün seslerin açık ve kapalı durumları vardır.

Ünlü sesler, ek formlarıyla (alofon) beraber incelenecek olursa, Türkçe'deki ünlülerin dizilimi (vowel quadrilateral) Şekil 4.1.'de ve bu özelliklerin sergilendiği örnek sözcükler Tablo 4.2'de gösterilmiştir [40]. Bu gösterimlerde ünlü sesleri temsil etmek için IPA [45] karakterleri kullanılmıştır.



Şekil 4.1. Türkiye Türkçe'sinin ünlü dörtgeni

Tablo 4.2. Türkçedeki ünlü sesler ve örnek sözcükler

Phoneme	Word	IPA
<b>ɒ</b>	anı - memento	<b>anı</b>
<b>a</b>	laf - talk	<b>laf</b>
<b>e</b>	elma - apple	<b>elma</b>
<b>ε</b>	dere - stream	<b>dere</b>
<b>i</b>	iğde - oleaster	<b>iyde</b>
<b>ɪ</b>	simit - bagel	<b>simit</b>
<b>ɨ</b>	ısı - heat	<b>isi</b>
<b>ɔ</b>	soru - question	<b>soru</b>
<b>o</b>	oğlak - goat	<b>oylak</b>
<b>ʊ</b>	kulak - ear	<b>kulak</b>
<b>u</b>	uğur - fortune	<b>uyur</b>
<b>œ</b>	örtü - blanket	<b>œrtü</b>
<b>ø</b>	öğren - learn	<b>øğren</b>
<b>ɥ</b>	ümit - hope	<b>ɥmit</b>
<b>y</b>	düğme - button	<b>dyğme</b>

Bunun yanı sıra Türkçedeki ünlüler uzunluklarına göre sınıflandırılmazlar. Her ünlünün uzun ve kısa şekli olur ve bir vokalin uzunluk bakımından normal, normalden uzun ve normalden kısa şekli olabilir [37]. Fakat Türkçede kısa ve normalden kısa vokal mevcut olup, uzun vokal yoktur. Normal Vokaller /a/,/e/,/i/,/o/,/ö/,/u/,/ü/ normalden kısa vokaller ise /ı/ vokalleridir [37].

Türkçedeki Arapça ve Farsça kelimelerde ise çok sayıda uzun vokal vardır. Bu yabancı uzun vokaller ise ses organlarının sun'î ve hususî bir gayreti ile söylenmektedir [5]. Bunun yanı sıra /ğ/ sesi, beraberinde kullanılan ünlülerin sürelerini uzatmaktadır. “Ağaç” kelimesinde /a/ sesinden sonra gelen /ğ/ , bu sesin uzun okunacağını göstermektedir. “Doğa” kelimesinde var olan ünlü kayması /o/ sesinden /a/ sesine doğru olmaktadır ve bu arada /o/ sesi yarım uzunluk kazanmaktadır [39].

Burada “Diftong” kavramından da bahsetmekte fayda vardır. İki vokalin bir hecede yan yana gelmesine diftong denir [37]. “Saadet” kelimesinde olduğu gibi iki adet /a/ harfi tek bir sesmiş gibi çıkarılır. Türkçede yabancı kökenli kelimeler dışında bu durum söz konusu değildir. Bu yüzden Türkçe kökenli kelimelerde diftong yoktur [39]. Ancak konuşma dilindeki kimi ünsüzlerin söyleyiş sırasında sesletilmemeleri nedeniyle, “tek sesmiş gibi çıkarılan ya da başlangıçta duyulan sesin bitişte yerini başka bir sese bırakması” olarak tanımlayabileceğimiz ünlü kaymalarının Türkçenin konuşa dilinde var olduğu sesyazarlarla yapılan ölçümler sonucu kanıtlanmıştır [33,39].

#### 4.4.2. Türkçedeki ünsüz sesler

Türkiye Türkçesinde ünsüz sesler, konsonant [37], sesdeşler [38] veya ünsüzler (consonant) [33,39] olmak üzere değişik terimlerle ifade edilebilmektedir. Yapılan incelemelerde görüldüğü üzere ünsüz seslerin hem sınıflandırılmasında hem de terminolojisinde konunun uzmanları tarafından bir uzlaşma sağlanamamıştır. Bu yüzden birkaç değişik kaynaktaki sınıflandırmalar incelenmiş ve nihayetinde

Ergenç'in [39] sınıflandırması ve terminolojisi baz alınıp diğer sınıflandırılmaldan da istifade ederek ses kümeleri ve sınıflandırmaları oluşturulmuştur.

Oluşumları sırasında ses yolunda (ses telleri, küçük dil, dil, damak, dişler ve dudaklarda) açılma, kapanma ya da daralma gibi bir engelle karşılaşılan seslerdir [34]. Ünsüzler diğer bir ifade ile engelli seslerdir. Çıkarılışları sırasında konuşma organlarının herhangi bir yerinde tutuklanırlar ve bunun sonucu olarak bir sürtünme ya da patlama biçiminde oluşurlar [39].

Ünsüz sesler çıkış biçimlerine, yerlerine ve ses tellerinin durumuna göre sınıflandırılırlar. Bu ana sınıflandırmada bir uzlaşma olsa da, alt sınıflandırmada farklılıklar görülmektedir. Bu alt sınıflandırmalar için [34,39,41]'deki sınıflandırmaları ayrı ayrı belirtmekte yarar görülmektedir. Sonuç olarak oluşturulacak ses kümesi ve diğer çıktılar için bu sınıflandırılmaldan istifade edilerek sentezleyici için uygun bir sınıflandırma yapılacaktır.

#### 4.4.3. Toklu'ya [34] göre ünsüz seslerin sınıflandırılması

Ünsüzler çıkış yerlerine ve çıkış biçimlerine göre iki bölüme ayrılır. Çıkış yerlerine göre ünsüzler aşağıda belirtilmiştir [42].

Çift Dudak ünsüzleri	: /b/, /m/, /p/
Diş – Dudak ünsüzleri	: /f/, /v/
Diş Eti ünsüzleri	: /d/, /l/, /n/, /r/, /s/, /t/, /z/
Diş Eti- damak ünsüzleri	: /c/, /ç/, /ş/, /j/
Ön damak ünsüzleri	: /g/, /k/, /y/
Art damak ünsüzleri	: /g/, /ğ/, /k/
Gırtlak Ünsüzü	: /h/

Ünsüzler çıkış biçimlerine göre de şu şekilde sıralanabilir:

#### 4.4.3.1. Süreksiz ünsüzler (Kapanma ünsüzleri)

Oluşumları sırasında akciğerden gelen hava bir engelle karşılaşır. Bu engeli aşmak için zorlama olunca, ses patlama biçiminde ortaya çıkar. Ses yolunda bir engelle karşılaşarak patlama biçiminde ortaya çıkan ünsüzlere süreksiz ünsüzler denir. /b/, /c/, /ç/, /d/, /g/, /k/, /p/ ve /t/ Türkçe'deki süreksiz ünsüzlerdir.

#### 4.4.3.2. Sürekli ünsüzler (Daralma ünsüzleri)

Ses yolunda herhangi bir engele rastlamadan çıkan ünsüzlere sürekli ünsüzler denir. Sürekli ünsüzlerin oluşumu sırasında ses yolunda çeşitli daralmalar olur. Ancak tam bir kapanma olmaz. Bu nedenle akciğerlerden gelen hava ses yolundan sızarak çıkar. Türkçedeki /f/, /ğ/, /h/, /j/, /l/, /m/, /n/, /r/, /s/, /ş/, /v/, /y/, ve /z/ sürekli ünsüz niteliği taşır.

#### 4.4.3.3. Ötümlü ve ötümsüz ünsüzler

Ses tellerinin titreşim durumuna göre ünsüzler ikiye ayrılır:

Ötümsüz (Sert) Ünsüzler : /ç/, /f/, /h/, /k/, /p/, /s/, /ş/, /t/

Ötümlü (Yumuşak) Ünsüzler : /b/, /c/, /d/, /g/, /ğ/, /j/, /l/, /m/, /n/, /r/, /v/, /y/, /z/

Bu bilgiler ışığında ünsüzlerin sınıflandırılması Toklu [34] tarafından Aksan [42] referans alınarak Tablo 4.3'teki gibi gösterilmiştir.



Tablo 4.3. Toklu'ya göre ünsüz seslerin sınıflandırılması (AKSAN, 1980:38)

			ÇIKIŞ YERLERİNE GÖRE						
			Çift-dudak	Diş-Dudak	Dişeti	Dişeti-Damak	Damak (Ön)	Art Damak	Gırtlak
ÇIKIŞ BİÇİMLERİNE GÖRE	Kapanma Ünsüzleri	Ötümlü	/p/		/t/		/k/	/k/	
		Ötümsüz					/g/	/g/	
	Kapanma-Daralma Ünsüzleri	Ötümlü				/ç/			
		Ötümsüz				/c/			
	Daralma Ünsüzleri	Ötümlü		/f/	/s/	/ş/			/h/
		Ötümsüz		/v/	/z/	/j/	/y/ (y.ünlü)	/ğ/	
	Burun Ünsüzleri	Ötümlü	/m/		/n/				
		Ötümsüz							
	Yan Ünsüzler	Ötümlü							
		Ötümsüz			/l/	/l/			
	Çarpmalı (Vurmalı) Ünsüzler	Ötümlü							
		Ötümsüz			/r/				

#### 4.4.4. Özsoy'a [41] göre ünsüz seslerin sınıflandırılması

Ünsüzlerin sesbilgisel özellikleri üç boyutta incelenir:

- 1) Oluşma noktası
- 2) Oluşma biçimi
- 3) Gırtlak durumu

##### 4.4.4.1. Oluşma noktasına göre:

Oluşturucular, alt ve üst oluşturucular olmak üzere ikiye ayrılır. Alt oluşturucular, üst oluşturuculara göre daha serbest olup seslerin oluşmasında daha etkindir. Alt ve Üst oluşturuculara göre seslerin oluşma noktaları Tablo 4.4 'te verilmiştir:

Tablo 4.4. Alt ve üst oluřturuculara gre nszler

Alt Oluřturucu	st Oluřturucu	Oluřma Noktası	Ses
Alt dudak	st dudak	ift- dudaksıl	[p]
Alt dudak	st diř	Dudaksıl- diřsil	[f]
Dil ucu	st diř	Dilucu- diřsil	[t]
Dil ucu	Diř yuvası	Diřyuvasıl	[s]
Dil palası	Diř yuvası - damak	Diřyuvasıl- damaksıl	[ř]
Dil palası	Damak	Damaksıl	[y]
Dil ardı	Art damak	Dil ardı – ard damaksıl	[k]
Dil ardı	Kk dil	Kk dilsil	[q]
	Boğaz	Boğazsıl	[?]
	Gırtlak	Gırtlaksıl	[h]

#### 4.4.4.2. Oluřma biimine gre:

nszler oluřma biimlerine gre; patlamalı, srtnmeli, patlamalı – srtnmeli, genizsil, kayıcı ve akıcı olmak zere 6 sınıfa ayrılır.

#### 4.4.4.3. Gırtlak durumuna gre:

Gırtlak durumu, hava akımı gırtlaktan geerken ses tellerinin durumunu belirtir. Ses tellerinin durumlarına gre tml, tmsz ve fısıltılı sesler oluřturulur.

Bu bilgiler ışığında zsoy'a gre nsz sınıflandırılması Tablo 4.5 te verilmiřtir.

Tablo 4.5. Özsoy'a göre ünsüz sınıflandırması [41]

		Du dak sıl	Diş sil / duda ksıl	Diş-sil	Diş eti- dişyuvasıl	Ön damaksı l	Damaks ıl	Art damak- sıl	Gırtl aksıl
Patlamalı	Ötümsüz	/p/		/t/			/k/	/k/	
	Ötümlü	/b/		/d/			/g/	/g/	
Sürtünmeli	Ötümsüz		/f/		/s/	/ş/			/h/
	Ötümlü		/v/		/z/	/j/			
Patlamalı/ sürtünmeli	Ötümsüz				/ç/				
	Ötümlü				/c/				
Genizsil	Ötümsüz								
	Ötümlü	/m/		/n/					
Akıcı	Ötümsüz								
	Ötümlü			/l/ /r/				/l/	
Kayıcı	Ötümsüz						/y/		
	Ötümlü								

#### 4.4.4.4. Hatipoğlu'na [36] göre ünsüz seslerin sınıflandırılması

Ünsüzleri sınıflandırırken genellikle üç ölçüt kullanılır: ses tellerinin titreşimi, eklemleme biçimi ve eklemleme noktası.

Söylenişleri sırasında ses telleri titreşmeyen ünsüzlere ötümsüz/titreşimsiz/sedasız sesler denir.

Eklemleme noktasına göre;

1. Çift dudaksıl (bilabial)
2. Dudaksıl – dişsil (labio -dental)
3. Dişsil –dişyuvasıl (denti -alveolar)
4. Dişyuvasıl-damaksıl (alveo - palatal)
5. Damaksıl (palatal)

6. Art damaksıl (velar)

7. Gırtlaksıl (glottal)

olmak üzere altı kategoride,

Eklemlene biçimine göre de;

1. Kapanma (stops)

2. Sürtünmeli (fricatives)

3. Kapanmalı - sürtünmeli (affricatives)

4. Genizsi (nasals)

5. Daralmalı (approximants)

6. Sızmalı (lateral aproximant)

7. Çarpmalı (flap / tap)

olmak üzere yedi kategoride sınıflandırılabilir.

Tüm bu bilgiler ışığında Hatipoğlu'na göre ünsüz sınıflandırılması Tablo 4.6.'da gösterilmiştir.

Tablo 4.6. Hatipoğlu'na göre ünsüz sınıflandırması [36]

Türkiye Türkçesinin Ünsüzleri								
		Çift dudak sıl (bilabial)	Dudaksıl – dişsil (labio-dental)	Dişsil – dişyuvasıl (denti-alveolar)	Dişyuvasıl –damaksıl (alveo-palatal)	Damaksıl (palatal)	Art damaksıl (velar)	Gırtlaksıl (glottal)
Kapanma (stops)	Ötümsüz	/p/		/t/		/k/	/k/	
	Ötümlü	/b/		/d/		/g/	/g/	
Sürtünmeli (Fricatives)	Ötümsüz		/f/	/s/	/ş/			/h/
	Ötümlü		/v/	/z/	/j/			
Kapanma-Sürtünmeli (Affricatives)	Ötümsüz				/ç/			
	Ötümlü				/c/			
Genizsi (Nasals)	Ötümlü	/m/		/n/				
Daralmalı (Approximants)	Ötümlü					/y/		
Sızmalı (Lateral approximant)	Ötümlü			/l/			/l/	
Çarpmalı (Flap / tap)	Ötümlü			/r/				

#### 4.4.4.5. Ergenç'e [39] göre ünsüz seslerin sınıflandırılması

Konuşma seslerinin ikinci büyük kümesini oluşturan ünsüzler, engelli seslerdir. Çıkarılışları sırasında konuşma organlarının herhangi bir yerinde tutuklanırlar ve bunun sonucu olarak bir sürtünme ya da patlama biçiminde oluşurlar. Ünsüzler çıkış biçimlerine, yerlerine ve ses tellerinin durumuna göre sınıflandırılabilir.

Çıkış Biçimlerine göre:

1. Patlamalı ünsüzler (Plosive) : /b/,/d/,/g/,/p/,/t/,/k/
2. Geniz ünsüzleri (Nasal) : /m/,/n/
3. Çarpmalı ünsüzler (Rolled) : /r/
4. Yan daralmalı ünsüzler (Lateral) : /l/
5. Sürtünücü ünsüzler (Fricative) : /c/,/ç/,/f/,/h/,/j/,/s/,/ş/,/v/,/y/,/z/

Çıkış Yerlerine Göre:

1. Çift dudak ünsüzleri (Bilabial) : /b/,/p/,/m/
2. Dudak – diş ünsüzleri (Labiodental) : /f/,/v/
3. Dilucu- dışardı ünsüzleri (Dental) : /d/,/t/
4. Dilucu – dişeti ünsüzleri (Palato -alveolar) : /n/,/r/,/s/,/z/
5. Dil – öndamak ünsüzleri (Alveo- palatal) : /c/,/ç/,/j/,/ş/,/y/
6. Dilucu – öndamak ünsüzleri (Apical /palatal) : /l/
7. Dil- art damak ünsüzleri (Velar) : /k/,/g/
8. Gırtlak Ünsüzleri (glottal) : /h/

Ses tellerinin titreşimine göre:

1. Ötümlü ünsüzler (Voiced) : /b/,/c/,/d/,/g/,/j/,/l/,/m/,/n/,/r/,/v/,/y/,/z/
2. Ötümsüz ünsüzler (Voiceless) : /ç/,/f/,/h/,/k/,/p/,/s/,/ş/,/t/

Yukarıda farklı referanslardaki sınıflandırılmalarda görüleceği üzere özellikle oluşma yerleri ve oluşma biçimlerine göre sınıflandırma farklılık göstermektedir. Sınıflandırılmalar incelenecek olursa gerek sınıf sayısının, gerek sınıf terimlerinin değiştiği görülmektedir. Tüm bu sınıflandırmalar konunun uzmanlarınca yapılmış olup, her bir sınıflandırmanın kendi açısından haklılık payı bulunmaktadır.

\* ”/ /” içinde gösterilen harfler, o harfe karşılık gelen sesleri ifade etmektedir.

#### 4.5. Harf – Ses Ayrımı

Her ne kadar Türkçe'nin yazıldığı gibi okunan bir dil olduğu söylene de meselenin derinlerine inildiği takdirde bu konunun öyle olmadığı görülecektir. Gerek seslerin oluşma yerleri, gerekse yanlış söyleyişte kelimelerde oluşacak anlam farkları bunu açıkça göstermektedir. Zaten konuşmakla iyi konuşmak arasındaki fark bu farkları göz önde bulundurmakla mümkündür.

Ölçünlü Türkçede bir harf birden fazla sese karşılık gelebilmektedir ve bu durumdaki seslere de “alofon” denir. Mevcut uluslar arası IPA [45] alfabesine göre de her sesin farklı bir gösterimi mevcuttur. Fakat okunabilirlik rahat olsun diye seslerin IPA gösterimleri yerine “/” arasında temsil edildikleri harfle gösterilmeleri yoluna gidilmiştir.

Daha önce de bahsettiğimiz üzere alfabemizdeki 29 harfe karşılık 44 fonemin bulunmaktadır. Yukarıdaki sınıflandırmalarda da bu husus kısmen görülmektedir. Bu mevzu Bölüm 4.5.1'den 4.5.2'e kadar incelenmiştir.

##### 4.5.1. /a/ Sesi

Geniş ve düz bir arkadil ünlüsü olan /a/ sesinin ağız boşluğunun önünde (predorsal) ve gerisinde (postdorsal) çıkartılan iki türü bulunmaktadır ve pradorsal /a/ sesi genellikle yabancı sözcüklerde bulunmaktadır [39]. Tablo 4.7'de /a/ sesinin durumları ile ilgili örnekler verilmiştir.

Tablo 4.7. /a/ Sesinin durumları

Tür	Örnek1	Örnek2	Örnek3
Predorsal /a/	Reklam	kâr	laf
Postdorsal /a/	Almak	kar	balık

#### 4.5.2. /e/ Sesi

Geniş, düz bir öndil ünlüsü olan /e/ sesinin üç türü olmakla beraber kapalı (closed) ve açık (open) /e/ sesleri en fazla kullanılan seslerdir [39]. Tablo 4.8'de /e/ sesinin durumları ile ilgili örnekler verilmiştir.

#### 4.5.3. /i/ Sesi

Kapalı- uzun (closed - long) ve açık – kısa (open -short) olmak üzere iki farklı /i/ sesi bulunmaktadır [39]. Tablo 4.9'de /i/ sesinin durumları ile ilgili örnekler verilmiştir.

#### 4.5.4. /o/ ve /ö/ Sesleri

Kısa- açık ve uzun-kapalı olmak üzere ikişer farklı /o/ ve /ö/ sesi bulunmaktadır [39]. Tablo 4.10'da /o/ ve /ö/ seslerinin durumları ile ilgili örnekler verilmiştir.

Tablo 4.8. /e/ Sesinin durumları

Tür	Örnek1	Örnek2	Örnek3
Kapalı (closed) /e/	Elma	seçim	ekşi
Açık (open) /e/	Dere	ses	içecek

Tablo 4.9. /i/ Sesinin durumları

Tür	Örnek1	Örnek2	Örnek3
Kapalı - uzun (closed -long) /i/	iplik	hatip	irfan
açık – kısa (open -short) /i/	iğne	kerime	itidal



Tablo 4.10. /o/ ve /ö/ Seslerinin durumları

Tür	Örnek1	Örnek2	Örnek3
Uzun – kapalı /o/	doğum	oğlak	Fotoğraf
Kısa – açık /o/	Kol	olgun	Ordu
Uzun – kapalı /ö/	Öğle	öğren	Başöğretmen
Kısa – açık /ö/	Örtü	köhne	Nöbet

#### 4.5.5. /u/ ve /ü/ Sesleri

Kısa- açık ve uzun-kapalı olmak üzere ikişer farklı /u/ ve /ü/ sesi bulunmaktadır [39]. Tablo 4.11'de /u/ ve /ü/ seslerinin durumları ile ilgili örnekler verilmiştir.

#### 4.5.6. /k/ Sesi

Ağız boşluğunun daha ön bölümünde üretilen /k/ sesi öndil ünlüleriyle ve predorsal /a/ sesiyle birlikte kullanılırken ağız boşluğunun daha gerisinde üretilen /k/ sesi arka dil ünlüleriyle seslendirilir [39]. Yani kalın ünlü tabir edilen arkadil ünlüleriyle beraber kullanılan /k/ sesi (arkadamak) ile ince ünlü diye tabir edilen ön dil ünlüleri ve predorsal /a/ sesi (“lâf” sözcüğünde olduğu gibi) ile birlikte kullanılan /k/ sesi (öndamak) birbirinden farklıdır. İngilizcedeki “q” ve “k” harfleriyle gösterilen sesler kuralların net olmasından dolayı tek bir “k” harfi ile gösterilmiştir. Tablo 4.12'de /k/ sesinin durumları ile ilgili örnekler verilmiştir.

Tablo 4.11. u/ ve /ü/ Seslerinin durumları

Tür	Örnek1	Örnek2	Örnek3
Uzun – kapalı /u/	tuğgeneral	uğur	Kûfi
Kısa - açık /u/	curcuna	ulak	Ufuk
Uzun – kapalı /ü/	düğme	üğrüm	Çobandüdüğü
Kısa - açık /ü/	Düş	ümit	Üzüntü

Tablo 4.12. /k/ Sesinin durumları

Tür	Örnek1	Örnek2	Örnek3
Kalın (arka damak) /k/	Kanı	kulluk	kat
İnce (öndamak) /k/	Kemal	küllük	kâğıt

#### 4.5.7. /g/ Sesi

/k/ sesi ile aynı duruma sahip olup, ön ve arkadamak sesleri birbirinden farklıdır. Tablo 4.13'de /g/ sesinin durumları ile ilgili örnekler verilmiştir.

#### 4.5.8. /l/ Sesi

Öndil ünlüleri ve predorsal /a/ ile beraber kullanılan /l/ sesi (öndamak) ve arkadil ünlüleriyle beraber kullanılan /l/ sesi (arkadamak) birbirinden farklıdır. Tablo 4.14'de /l/ sesinin durumları ile ilgili örnekler verilmiştir.

Tablo 4.13. /g/ Sesinin durumları

Tür	Örnek1	Örnek2	Örnek3
Kalın (arkadamak) /g/	Goril	galata	gurultu
İnce (öndamak) /g/	Gelin	gönül	gürültü

Tablo 4.14. /l/ Sesinin durumları

Tür	Örnek1	Örnek2	Örnek3
Kalın (arkadamak) /l/	kalın	albay	kurtuluş
İnce (öndamak) /l/	kilo	lale	karanfil

#### 4.5.9. /v/ Sesi

/a/ ve /u/ sesi arasındaki /v/ sesi yarı ünlü karakteri kazanır [39]. Bu durumda İngilizce ve diğer dillerde olan fakat Türkçe’de olmayan “w” harfini karşılar. Tablo 4.15'te /v/ sesinin durumları ile ilgili örnekler verilmiştir.

#### 4.5.10. /r/ Sesi

Bu sesin üç farklı durumu vardır. Önseste iken çok vuruşlu, içseste tek vuruşlu, sonseste ise sürtünücü karakterde olan farklı /r/ sesleri mevcuttur. Tablo 4.16'da /r/ sesinin durumları ile ilgili örnekler verilmiştir.

#### 4.5.11. /z/ Sesi

Sonseste ötümlü olan /z/ sesi ötümsüzleşir. Tablo 4.17'de /z/ sesinin durumları ile ilgili örnekler verilmiştir.

#### 4.5.12. /n/ Sesi

/k/, /g/, /y/ ve /j/ sesleri ile beraber bulunan /n/ sesi genizsileşir. Tablo 4.18'de /n/ sesinin durumları ile ilgili örnekler verilmiştir.

Tablo 4.15. /v/ Sesinin durumları

Tür	Örnek1	Örnek2	Örnek3
Yarı ünlü /v/	yavuz	tavuk	kavun
Ünsüz /v/	vatan	vurgun	tavla

Tablo 4.16. /r/ Sesinin durumları

Tür	Örnek1	Örnek2	Örnek3
Önses – çok vuruşlu /r/	roman	ruh	rüzgar
İç ses – tek vuruşlu /r/	ırak	orman	taşra
Son ses – sürtünücü /r/	bBir	memur	kindar

Tablo 4.17. /z/ Sesinin durumları

Tür	Örnek1	Örnek2	Örnek3
Ötümlü /z/	zorba	ziyaret	zümrüt
Ötümsüz /z/	kiraz	karpuz	öksüz

Tablo 4.18. /n/ Sesinin durumları

Tür	Örnek1	Örnek2	Örnek3	Örnek4
Genizsi /n/	Ankara	engin	Konya	anjın

#### 4.5.13. /ğ/ Olgusu

“ğ” harfi tek başına bir seslem olarak kabul edilmese de beraber kullanıldığı ünlüleri uzun okuması bakımından önemli görevlere sahiptir. /ğ/ söyleyişte yittiği durumlarda /ğ/’den önce gelen ünlü uzar [33].

/ğ/ sesinin durumlarını ayrı ayrı inceleyecek olursak: \*

1. a+ğ dizisini /a/, /u/ ve /ı/ sesleri izleyebilir. “ağam”, “ağustos”, “ağıt”
2. “ağustos”, “ağıt” kelimelerinde ünlü kayması oluşmuştur. “ağam” kelimesinde ise /a/ sesi uzamıştır [33].
3. o+ğ+[k] dizisinde /o/ ünlüsü uzundur.

4. o+ğ+[v] dizisinde dilimizdeki ünlü uyumuna uygun olarak vokal yerine /a/ veya /u/ sesleri gelebilir: “doğa” ve “oğul”. Bu kelimelerde iki ünlü yan yana kalsa da ünlü kaymasından söz edilemez [33].
5. u+ğ+[v] dizisinde vokal yerine yalnızca /u/ ünlüsü gelebilir. “uğultu” Örneğinde olduğu gibi /ğ/ nin etrafında ki ünlüler birleşmiş ve uzun bir /u/ sesi oluşmuştur [33].
6. e+ğ dizisini dilimizdeki ünlü uyumuna göre /e/ ve /i/ sesleri izleyebilir [33].
7. e+ğ+e dizisinde , örneğin “eğe” sözcüğünde olduğu gibi, dil tam bir devinim yapar ve buradaki ğ harfi /y/ sesine dönüşür [33].
8. e+ğ+i dizisinde , “eğitmen” örneğinde olduğu gibi, /e/ ünlüsünü izleyen /ğ/ söyleyişte yitirilmiş ve /e/ sesi /i/ sesine kaymıştır. Söyleyişte yitirilen ğ, /y/ sesine dönüşür ve bu ünsüzün çıkışında dil /i/ ünlüsünün aldığı biçime dönüştüğünden dolayı yarı ünlü niteliği kazanır [33].
9. e+ğ dizisini herhangi bir ünsüz izleyebilir. “eğmek” örneğinde olduğu gibi önseste bulunan /e/ sesi ğ nedeniyle uzar [33].
10. i+ğ dizisini ünlü olarak yalnızca /i/ ünlüsü izleyebilir. “yiğit” örneğinde görüleceği üzere /i/ ünlüleri birleşerek uzar [33].
11. i+ğ+[k] dizisinde konsonant yerine her ünsüz gelebilir. “iğne” Örneğinde olduğu gibi önseste bulunan /i/ kapalıdır ve onu izleyen ğ /y/ sesine dönüşür [33]. Fakat Ergenç’in verdiği örneklerde bu durumdaki /i/ sesi uzun okunmaktadır [39].
12. ö+ğ+[v] dizisinde vokal yerine /e/ ve /ü/ sesleri gelebilir. “öğüt” örneğinde olduğu gibi uzama olmaz. Ünlü kayması oluşur [33].
13. ö+ğ+[k] dizisinde ise konsonant yerine akıcı ünsüzler /l/ ve /r/ ile geniz ünsüzlerinden çift dudak /m/ sesi gelebilir. “öğretmen” örneğinde olduğu gibi bu durumda yalın halde söylenen ünlü uzunluğundan daha uzun değilse de aynı sözcükteki diğer ünlülerden daha uzundur [33].
14. ü+ğ+[v] dizisinde vokal yerine yalnızca /ü/ ünlüsü gelebilir. “düğün” örneğinde olduğu gibi /ğ/’den önceki /ü/ sonraki /u/ sesine ulanarak uzar. Söyleyişte aradaki

/ğ/ ön dil ünlülerinin hemen hepsinde olduğu gibi /y/ ünsüzüne dönüşür ve bu ses yarı ünlü olarak kabul edildiğinden ünlü kayması görülür [33].

15. [k]+ü+ğ dizisinde ise “düğme” örneğinde olduğu gibi /ü/ sesi “dügün” deki /ü/ sesi kadar olmasa da uzun okunur. Bu yüzden yarı uzundur denilebilir[33].
16. /ı/ sesi /ğ/ eşliğinde diğer ünlülere göre daha dikkate değer bir durum gösterir. /ğ/ Sonseste olduğunda “sığ” örneğinde olduğu gibi /ı/ sesi yarı uzundur.
17. [k]+[v]+ğ+[v] dizisinde ikinci vokal yerine yalnızca /ı/ sesi gelebilir. “yığılmak” ya da “yığıl” örneğinde olduğu gibi. Bu sözcüklerde dil, /ı/ ünlüsü çıkışında /o/ ünlüsünün oradan da /u/ ünlüsünün çıkış biçimine giderek kaymıştır. Bu duruma göre üçlü ünlü kayması olmuştur denebilir [33].

Tüm bu durumlara göre /ğ/ sesi kullanıldığı zamanlarda ya ünlü kayması olur, ya beraber kullanıldığı ünlüyü uzatır ya da /y/ sesine dönüşür ki bu durumda da yarı ünlü karakteri kazandığından sözde ünlü kayması oluşur [39]. Tablo 4.19'da /ğ/ sesinin durumları ile ilgili örnekler verilmiştir.

Tablo 4.19. /ğ/ Etkileri

/ğ/ etkisi	Örnek1	Örnek2	Örnek3
Ünlü Uzaması	uğur	dağ	yağmur
Ünlü Kayması	ağıt	doğa	öge
Sözde Ünlü Kayması (/y/ sesi)	eğitim	eğlence	eğe

\* /ğ/ sesi incelemesinde [v] ünlü sesleri (vokal), [k] ise ünsüz sesleri (konsonant) temsil etmektedir.

Bazı kaynaklarda ötümlü ve ötümsüz sesler sedalı ve sedasız [37], veya titreşimli/sedalı ve titreşimsiz/sedasız [43], sert/sessiz ve yumuşak/sesli [44] olarak adlandırılmıştır.

Tüm bu bilgiler göz önünde bulundurularak oluşturulan Türkçe Ses Kümesi ve ses kümesiinde kullanılan parametrelerin açıklamaları Tablo 4.20 ve Tablo 4.21'de verilmiştir.

Tablo 4.20. Türkçe Ses Kümesinde kullanılan parametreler

Ünlü veya ünsüz	Ünlü	Ünsüz					
vc	+	-					
Ünlü uzunluğu	kısa	uzun	Daha uzun				
vlnɡ	<b>s</b>	<b>l</b>	<b>d</b>	<b>0</b>			
Ünlü genişliği	dar	orta	geniş				
vheight	<b>1</b>	<b>2</b>	<b>3</b>	<b>0</b>			
Ünlü yeri	Ön (ince)	orta	Arka (kalın)				
vfront	<b>1</b>	<b>2</b>	<b>3</b>	<b>0</b>			
Ünlü (dudak) yuvarlaklığı	yuvarlak	düz					
vrnd	+	-	<b>0</b>				
Ünsüz tipi	patlamalı	sürtünücü	yarı - kapanmalı	genizsi	yan daralmalı	daralmalı	
ctype	<b>s</b>	<b>f</b>	<b>a</b>	<b>n</b>	<b>l</b>	<b>r</b>	<b>0</b>
Çıkış noktası	dudak	dilucu	damaksıl (öndamak)	dudak dış	dış	dil-artdamak	gırtlak
cplace	<b>l</b>	<b>a</b>	<b>p</b>	<b>b</b>	<b>d</b>	<b>v</b>	<b>ɡ</b>
Consonant voicing	ötümlü	ötümsüz					
cvox	+	-					

Tablo 4.21. Türkçe ses kümesi

SES	Ses İşareti	v/c	v/ing	vheight	vfront	vrnd	ctype	cplace	cvox	Örnek	Açıklama
/a/	a	+	L	3	3	-	0	0	0	anı	postdorsal
	a1	+	L	3	3	-	0	0	0	laf	predorsal
/e/	e	+	L	3	1	-	0	0	0	elma	kapalı
	eh	+	L	3	1	-	0	0	0	tarife	açık
/i/	ih	+	s	1	2	-	0	0	0	ısı	ortadil
/i/	i	+	L	1	1	-	0	0	0	simit	kapalı
	i	+	d	1	1	-	0	0	0	iğde	açık
/o/	o	+	L	3	3	+	0	0	0	soru	açık
	oo	+	d	3	3	+	0	0	0	oğlak	kapalı
/ö/	oe	+	L	3	1	+	0	0	0	örtü	açık
	oe	+	d	3	1	+	0	0	0	öğren	kapalı
/u/	u	+	L	1	3	+	0	0	0	kulak	açık
	uu	+	d	1	3	+	0	0	0	uğra	kapalı
/ü/	ue	+	L	1	1	+	0	0	0	ümit	açık
	uee	+	d	1	1	+	0	0	0	düğme	kapalı
/b/	b	-	0	0	0	0	s	l	+		
/c/	c	-	0	0	0	0	a	p	+		
/ç/	ch	-	0	0	0	0	a	p	-		
/d/	d	-	0	0	0	0	s	d	+		
/f/	f	-	0	0	0	0	f	b	-		
/g/	g	-	0	0	0	0	s	v	+	karga	artdamak
	gh	-	0	0	0	0	s	v	+	genç	öndamak
/ğ/	gy	-	0	0	0	0	f	v	+		
/h/	h	-	0	0	0	0	f	g	-		
/j/	j	-	0	0	0	0	f	p	-		
/k/	k	-	0	0	0	0	s	v	-	kalas	Artdamak
	q	-	0	0	0	0	s	v	-	kemal	Öndamak
/l/	l	-	0	0	0	0	l	a	+		
	ly	-	0	0	0	0	l	a	+		
/m/	m	-	0	0	0	0	n	l	+		
/n/	n	-	0	0	0	0	n	a	+	nalan	
	ng	-	0	0	0	0	n	v	+	engin	genizsi
/p/	p	-	0	0	0	0	s	l	-		
/r/	r	-	0	0	0	0	r	a	+	ırmak	tekli
	rr	-	0	0	0	0	r	a	+	raf	çoklu
	rh	-	0	0	0	0	f	a	-	bir	sürtünücü
/s/	s	-	0	0	0	0	f	a	-		
/ş/	sh	-	0	0	0	0	f	p	-		
/t/	t	-	0	0	0	0	s	d	-		
/v/	v	-	0	0	0	0	f	b	+	var	ünsüz
	w	-	0	0	0	0	r	b	+	yavuz	yarı ünlü
/y/	y	-	0	0	0	0	r	p	+		
/z/	z	-	0	0	0	0	f	a	+	azık	ötümlü
	zh	-	0	0	0	0	f	a	-	yoz	ötümsüz



#### 4.6. Türkiye Türkçe'sinin Bürünsel Özellikleri

Dillerde sesbirimsel çizginin dışında kalan, parça (segment) özelliği taşımayan, parçalar üstü (supra segmental) birimler de vardır. Bu parçalar üstü birimler, sesin yeğnliğinden, yüksekliğinden, süresinden kaynaklanır ve bürün olguları (prozodi) olarak adlandırılır [39].

Bürün olguları her dilde dilbilimsel açıdan anlam ayırt edici bir işlev yerine getirmez. Çünkü dilbilimsel nitelikleri belirleyen, fiziksel olgular değil, karşıtlıklar ve ayrılıklarla bunların yol açtığı bildirişim değeridir. Bu nedenle, bürünsel olguların tümü de her dilde anlam birimleri, sözceleri birbirinden ayırt etmek için kullanılmaz. Bürün olguları, işlevsel bir değer taşıdığında bürün birim (prosodeme) adını alır[46]. Örneğin, parçalar üstü ses birimlerden vurgunun bürün birim olarak kabul edilebilmesi için bulunduğu dilde ayırt edici işlev üstlenmesi, bunun için de değişiklik göstermesi gerekir.Vurgu, Türkçe'nin yanı sıra İspanyolca, İtalyanca, İngilizce gibi dillerde bürün birim niteliği taşımaktadır [47].

Vurgunun yanı sıra süre, kavşak ve durak, ton ve ezgi de Türkçede ayırt edici işlev üstlendiğinden bunları bürün birim olarak kabul etmek mümkündür. Bölüm 4.6.1'den, 4.6.5'e kadar bu bürün birimlere değinilecektir.

##### 4.6.1. Süre

Ünlü sesbirimlerin çıkarılış süreleri (duration, length) Türkçede anlam ayırt edici güçtedir. Türkçe ünlüler asıl uzunluklarını kaybettiklerinden, uzun ünlüler ya yabancı kökenli sözcüklerde ya da ses yitimiyle oluşan özel durumlarda karşımıza çıkmaktadır [39]. Süre bilgisi ile ilgili örnekler Tablo 4.22'de gösterilmiştir.

Tablo 4.22. Süre Bilgisi

Süre	Örnek1	Örnek2	Örnek3
Uzun	düğün (dü:n)	katil (ka:til)	dağ (da: )
Kısa	dün (dün)	katil (katil)	da (da)

#### 4.6.2. Kavşak ve durak

Kavşak (juncture), ulamalardan kaynaklanacak anlam kaymalarını önleyebilmek için sonesinde ünsüz bulunan bir kelime ile ilk sesinde ünlü bulunan başka bir kelimenin söylenişi arasında verilen kısa bir aradır [33,39] . Bu durumlarla ilgili örnekler Tablo 4.23'te verilmiştir.

Türkçede kavşakları belirleyen özellikler kavşak noktasındaki seslerin soluklu / soluksuz, ötümlü/ötümsüz, vurucu/akıcı olması ve ünlü ile başlayan seslemlerden önce gırtlaksız patlamalı sesin bulunmasıdır [41].

Durak (pause) ise bir tümce ya da sözce içinde bulunan birden çok bilgi öbeği arasında verilen kısa aralardır. Yazılı dilde durakların belirlenmesi amacıyla noktalama işaretlerinden faydalanılır. Duraklar gereği gibi kullanılmadığı durumlarda yanlış anlamalara neden olacak ölçüde önemli ve anlam ayırıcı özellik taşımaktadırlar [39].Durak bilgisi ile ilgili örnekler de Tablo 4.24'te verilmiştir.

Tablo 4.23. Kavşak Bilgisi

Durum1	Durum2
O balkona çıkmış.	O balkon açılmış.
Adam yatak almış.	Adam yatakalmış.
Kesiti getir.	Kes iti getir.

Tablo 4.24. Durak Bilgisi

Durum1	Durum2
İzinsiz / inşaata giren iki çocuk yaralandı.	İzinsiz inşaata / giren iki çocuk yaralandı.
Oku/ Ahmet gibi yaramaz /olma.	Oku/ Ahmet gibi/ yaramaz olma.

### 4.6.3. Ton

Ton (tone), bir seslemdeki frekans yüksekliği veya düşüklüğü, yani bir seslemin tiz ya da pes olarak sesletilmesi olarak tanımlanır [39]. Ton, Çince, Japonca, İsveççe, Norveççe gibi dillerde sözcüklere yeni anlamlar yüklerken [45], Türkçede ezgi birimine bağlı olarak genellikle tek sözcükten oluşan bildirilerde anlam ayırıcı bir özellik taşır.

Ton kullanılması tamamıyla dillere ve kişilere göre değişiklik gösterebilir. Tonlama olayı Türkçe için zayıfken, İngilizce için çok zayıf, İspanyolca ve İtalyanca içinse önemlidir. Tablo 4.25'te ton bilgisi ile ilgili örnekler verilmiştir.

### 4.6.4. Ezgi

Ezgi (intonation), bir konuşma zincirindeki seslem, biçimbirim ve sözcükleri kapsayan ton değişimlerinin tümü olarak adlandırılır [39].

Tablo 4.25. Ton Bilgisi

Durum1	Durum2
Efendim (↑) (anlamadım manasında)	E (↑) fendim (çağrıya yanıt)
Ha: (↑) (anlamadım)	Ha: (↓) (anladım)

Konuşmanın ezgisi de konuşan kişiye ve konuşulan dile göre tamamen farklılık göstermektedir. Bu sebepten dolayı ezgiyi kesin kurallara oturtmak zordur. Tüm bu zorluklara karşın her dil için ezgi sistemi oluşturulmaya çalışılmaktadır. Türkçenin ezgi analizleri için Selen'in [48] kitabına bakılabilir. Ezginin kullanımı tamamen dile ve konuşmacıya ve koşullara bağlı olduğundan, her dil için belirlenmiş kimi temel ezgi biçimleri dışında evrensel nitelikli kurallar koymak olanaksız olsa da üç tür ezgiden söz edilebilir.

1. Biten ezgi: Cümlelerin bittiğini ve iletilmek istenen bildirinin sona erdiğini dinleyiciye iletmek işleviyle ses tonunun cümlelerin sonunda düşmesiyle ortaya çıkar.
2. Süren Ezgi: Genellikle yan cümlelerden ve sıralı cümlelerden oluşmuş bildirilerde dinleyiciye bildirim sürecini iletmek üzere ses tonunun ezgi doruğuyla aynı düzeyde kaldığı ya da bir iki perde yükseldiği durumlarda ortaya çıkar.
3. Soru Ezgisi: Dinleyiciden herhangi bir konuda bilgi ya da yanıt istendiğinde, ses tonunun cümle sonunda yükselmesiyle ortaya çıkar. Türkçede soru eki olan [-mı] vurgu almadığı için ezgi doruğu bir önceki seslemin üzerinde kalır.

Ezgi türlerine göre örnekler Tablo 4.26. 'da verilmiştir.

Bunlardan başka bir de odak ezgisi mevcuttur. Odak (focus), yüksek ezgi ile vurgunun aynı öge üzerinde gerçekleştirilmesi ile belirtilir. Böylelikle odak cümleleri yansız cümlelerden vurgu, ezgi ve örüntüleri bakımından farklılık sergiler [9]. Tablo 4.27'de bu durum ayrıntısıyla verilmiştir. 1,2,3 üstsel değerleri vurgu derecelerini belirtir.

Tablo 4.26. Ezgi Türleri

Ezgi Türü	Örnek
Biten Ezgi	Bunu beğenmedim. (↓)
	Çocuklar sinemaya gittiler. (↓)
Süren Ezgi	Geldim, (↑)(→) gördüm, (↑)(→) yendim. (↓)
	Bir sigara içip (↑)(→) hemen geleceğim. (↓)
Soru Ezgisi	Geliyor (↑) mu? (↓)
	Yazacak (↑) mısınız? (↓)

Tablo 4.27. Odak Ezgisi

Odak Ezgisi	Örnek	Durum
Yansız	2(Biz) 3bu akşam evde oturacağız1.	
Özne Odaklı	3(Biz) 2bu akşam evde oturacağız1.	Onlar değil
Zaman odaklı	2(Biz) bu ak3şam evde oturacağız1.	Sabah değil
Yer odaklı	2(Biz) bu akşam ev3de oturacağız1.	Komşuda değil

Odaklanma ezginin en önemli işlevlerinden biridir ve konuşmacının, bildirim işlevi sırasında dinleyicinin dikkatini çekmek istediği bilgi öbekleri üzerinde yaptığı ton değişikliği olarak ta adlandırılabilir ve çeşitli biçimlerde gerçekleştirilebilir [39]. Bunlardan biri yukarıda belirtildiği gibi sözdiziminde değişiklik yapılmadan belirtmek istenilen öge vurgulu bir şekilde söylenir. Diğerinde ise yazı dilinde yapılacak değişiklikle odaklanılacak öge, yüklem önüne getirilir.

#### 4.6.5. Vurgu

Sözcük vurgusu (accent), bir konuşmanın akışı içinde kullanılan sözcüğün bir seslemine öne çıkarmak için diğerlerine oranla daha baskılı, daha soluklu söylemek biçiminde tanımlanabilir [39,35].

Türkçede vurgu, birincil ve zayıf vurgu olmak üzere iki vurgu düzeyinde gerçekleşir. Sözcük düzeyinde, kök sözcüklerin vurgu düzeneği ile bileşik sözcük ve tamlama yapılarının vurgu düzeneği birbirinden farklı özellikler sergiler. Çok seslemlili kök sözcüklerde iki ayrı vurgu derecesi bulunurken, bileşik sözcük veya tamlama yapılarında üçüncül vurgu düzeyi bulunmaktadır [41]. Fakat zayıf vurgu genellikle göz önünde bulundurulmaz ve Türkçede sözcüklerde vurgu, karşıt bir durum söz konusu değilse, son seslemdedir [49]. Örneğin “kadar” sözcüğünde sözcüğün ilk seslemi olan –ka zayıf vurgu, ikinci seslemi olan –dar ise birincil vurgu alır. Buradaki zayıf vurgudan çoğu kaynakta bahsedilmez ve sadece son seslemdeki vurgu sözcük vurgusu olarak adlandırılır. Bu konunun çerçevesi bu raporu aşacağından, geniş bilgi için Özsoy’un [41] kitabına bakılabilir. Bunun yanı sıra ikinci seslemleri vurgulu olan “kara” ve “tahta” sözcüklerinden birleştirme yoluyla elde edilen “karatahta” sözcüğünde üçüncül vurgu ikinci sözcük olan “tahta” sözcüğünün son sesleminde görülmektedir.

Türkçe sözcüklerde vurgu durumu incelenecek olursa:

Türkçe ad ya da eylem köklerinde vurgu her ne kadar son seslemdeyse de üzerine vurgu almayan bir ekle kesilmedikçe sözcüğün son ekine kadar gider [35,39].

<b><u>Onur</u></b>	<b><u>Okul</u></b>
<b><u>Onurlan</u></b>	<b><u>Okullar</u></b>
<b><u>Onurlandır</u></b>	<b><u>Okullardan</u></b>
<b><u>Onurlandırıl</u></b>	
<b><u>Onurlandırılmış</u></b>	

Vurgu eylem veya ad kökünün son sesleminde değilse yer değiştirmez.

<b><u>Teyze</u></b>
<b><u>Teyzemin</u></b>
<b><u>Teyzemlerden</u></b>

Eylem ya da ek köküne getirilen eklerden birisi, üzerine vurguyu çekmeyen bir ekse, sözcük vurgusu bu ekten önceki seslemde kalır [39].

<b>git</b>	Gidile <u>mi</u> yorsa
<b>yaz</b>	Yaz <u>ı</u> yor

Türkçede vurguyu üzerine almayan son ekler ve seslemler Tablo 4.28'de verilmiştir [41].

Tablo 4.28. Türkçede Üzerine Vurguyu Almayan Ekler

Son Ek Tipi	Son Ekler	Örnek	
Emir Eki	-In / -InIz	<b>Gel</b> -in	<b>Gel</b> -iniz
Olumsuzluk Eki	-mA	<b>Bak</b> -ma	Kon <u>u</u> s-ma
Ulaş	-ken	Koş+ar+ken	Koş <u>ar</u> ken
Parçacık Tipi	Parçacık Eki	Örnek	
Eşlik	-(y)lA	Ayşe + ile	Ayş <u>e</u> yle
Soru	mI	Ayşe + mi	Ayş <u>e</u> mi
Odak	Da	Ayşe +de	Ayş <u>e</u> de
Görünüş	-(y)mIş	Ayşe+ymiş	Ayş <u>e</u> ymiş
Koşul	-(y)sA	Ayşe ise	Ayş <u>e</u> yse
Yer	-rA	Bu+ra	<b>Bu</b> rası

Türkçe’de üzerine vurgu almayan kimi eklerin vurgu alabilen sesteşleri bulunmaktadır. Örneğin, olumsuzluk eki –mA üzerine vurgu almazken, fiilden isim yapım eki olan sesteşi vurgulanır. Vurgunun anlam ayırıcı işlevi bu şekildeki sesteş kelimelerde daha iyi ortaya çıkmaktadır [39].

<b><u>Ta</u>stı</b>	Ta <b><u>stı</u></b>
<b><u>Gü</u>ldü</b>	Gü <b><u>ldü</u></b>
<b><u>Va</u>rdı</b>	Va <b><u>rdı</u></b>
Gü <b><u>lü</u>m</b> se	Gü <b><u>lü</u>m</b> se
<b><u>Dü</u>ştü</b>	Dü <b><u>ştü</u></b>
Ko <b><u>nu</u>ş</b> ma	Ko <b><u>nu</u>ş</b> ma

Aynı durum sesteş kelimelerden birisinin yer adı olması durumunda da geçerlidir. Yer adlarında vurgu ilk sesleme geçer [39]. İki den fazla seslemi olan yer adlarında ise vurgu yerleştirmesi sözcüğün seslem yapısına bağlı olarak gerçekleşir. Bu sözcüklerde sözcüğün son sesleminden önce gelen seslemlerin açık ya da kapalı olması vurgunun hangi seslem üzerinde gerçekleşeceğini belirler [41].

Cins ad	Yer adı	Cins ad	Yer adı
bo <b><u>drum</u></b>	<b><u>Bod</u>rum</b>	armu <b><u>tl</u></b>	Ar <b><u>mut</u></b> lu
to <b><u>kat</u></b>	<b><u>Tok</u>at</b>	ova <b><u>çık</u></b>	Ov <b><u>a</u></b> çık
or <b><u>du</u></b>	<b><u>Or</u>du</b>	köy <b><u>ç</u>ğ</b> iz	<b><u>Kö</u>y</b> çğiz

#### 4.7. Sonuç

Mevcut konuşma teknolojilerinin hemen hepsi doğası gereği dil teknolojilerinden destek almaktadır. Konuşma sentezleme uygulamalarının yanı sıra, konuşma ve konuşmacı tanıma uygulamalarında da mevcut dil modellerinden faydalanılarak başarımlarının artırılmasına gidilmektedir.



Sentezleme teknolojilerindeki temel birimler konuşmayı üreten seslerdir ve bu sesler her dil için farklılık göstermektedir. Neticede insandaki ses üretme mekanizması elliden fazla değişik konuşma sesi üretiyor olsa da her dil bu seslerden kendine has olanları içermektedir. Hatta belli bir süre sonra insanın hançere yapısı o hale gelir ki farklı sesleri çıkarmaya çalışırken doğru bir şekilde seslendiremez. Yabancı bir insanın Türkçe konuşurken konuşmasının farklı gelmesinin en büyük sebebi de budur. Aynı durum sentezleyicilerde de kendini göstermektedir. Örneğin İngilizce için gerçekleştirilmiş bir sentezleyicide Türkçe bir metni sentezlemeye çalışılırsa, konuşma farklılığı burada da görülecektir.

Bu bölümde fonetik alt yapının bu kadar irdelenmesinin sebebi Türkçenin ses yapısını mümkün olduğunca iyi bir şekilde modelleyebilmektir. Bu modellemenin doğru olması, zincirleme olarak toplanacak veritabanının daha sağlıklı olmasını, doğal dil işleme işlemlerinde girdilerin daha sağlıklı olmasını ve nihayetinde de sentezlenecek konuşmanın kalitesini etkilemektedir. Ayrıca, Türkçe ses yapısının iyi modellenmesi, geliştirilecek MKS sisteminin başarımı için de büyük önem arz etmektedir.

Türkçenin fonetiği her ne kadar analiz edilmeye çalışılmış olsa da geliştirilecek sentezleyicide bu bilgilerin kullanılması bu tezin kapsamını çok aşmaktadır. Bu sebepten gerçekleştirilecek sentezleyici sisteminde ses- harf ayırımına dikkat edilmemiştir. Bu çalışma neticesinde sentezleme için gerekli olan fonetik alt yapı oluşturulmaya çalışılmıştır. Konuşma sentezleme teknolojilerinde kullanılan harften sese çevrim kuralları bu rapordaki bilgiler ve çıktılar doğrultusunda gerçekleştirilmiştir ve “Türkçenin Söyleyiş Kuralları” raporunda bu kurallar verilmiştir. Bunu yanı sıra sentezlenecek konuşmanın doğallığını belirleyecek olan bürünsel niteliklerle ilgili altyapı da bu rapor kapsamında belirlenmiştir

## **BÖLÜM 5. KONUŞMA SENTEZLEME**

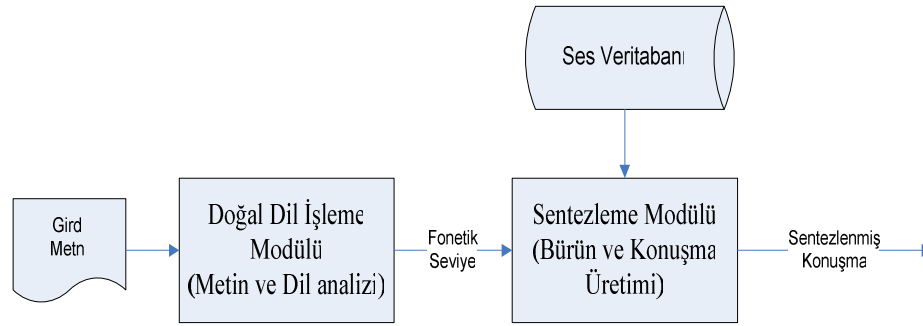
### **5.1. Genel**

Konuşma sentezleme insan sesinin yapay olarak üretilmesidir. Bu amaçla kullanılan bilgisayar sistemine konuşma sentezleyici denir. Bir sentezleme sistemi yazılımsal ve donanımsal olarak gerçekleştirilebilir. Bu bölümde konuşma sentezleyicinin temel yapısı hakkında bilgi verilecek ve geliştirilen Türkçe konuşma sentezleme sisteminde gerçekleştirilen adımlar anlatılacaktır.

### **5.2. MKS Sisteminin Aşamaları**

Bir MKS sistemi temelde üç farklı aşamadan oluşur:

1. Ses Veritabanının Hazırlanması
2. Doğal Dil İşleme Modülü (Metin ve Dil Analizi)
3. Sentezleme Modülü (Bürün ve Konuşma Üretimi)



Şekil 5.1. Sadeleştirilmiş MKS Sentezleme süreci [50]

### 5.2.1. Ses veritabanının hazırlanması

Metinden Konuşma sentezleme sistemlerinde sistem tasarımı hazırlanacak ses veritabanına göre gerçekleştirilmelidir. Formant sentezleme veya artikülasyon bazlı sistemlerde herhangi bir ses veritabanına ihtiyaç yoktur. Eklemeli yöntemlerde ise uygulanacak tekniğe göre bir bütüncü tasarlanmalı ve ses veritabanı bu bütüncüye göre oluşturulmalıdır. Sistemin genel kalitesi ses veritabanının hazırlanması ile doğrudan ilişkilidir. Sentezleme, hedef metni oluşturacak parçaların ses veritabanından seçilip birbirine eklenmesi suretiyle gerçekleştirilir.

### 5.2.2. Doğal dil işleme modülü (Metin ve dil analizi)

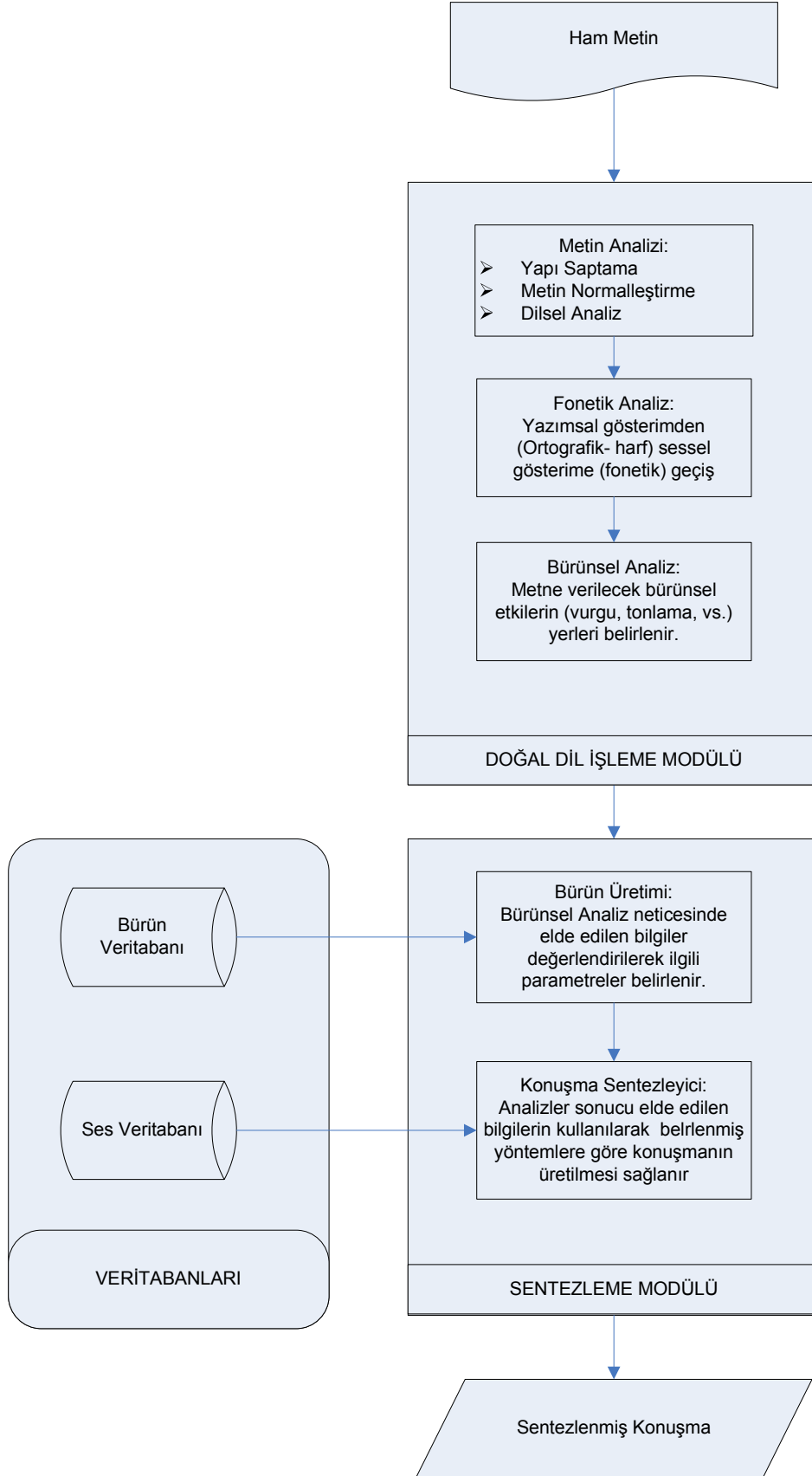
Bu aşamada metinden daha fazla enformasyon elde edilmeye çalışılır. Metin normalleştirilmesi, metnin yapısının belirlenmesi, metnin fonetik açılımının elde edilmesi, süre, vurgu ve tonlama bilgilerinin elde edilmesi hep bu analizler neticesinde elde edilir. Kısacası Doğal Dil İşlemenin aşaması olarak ta öngörülebilir. Burada önemli olan konuşmanın üretileceği dilin kurallarının ve akustik özelliklerinin iyice belirlenip, konuşma üretmeden önce gerekli parametrelerin doğru bir şekilde belirlenmesidir.

### 5.2.3. Sentezleme modülü (Bürün ve konuşma üretimi)

Bu aşamada metin ve dil analizinden elde edilen bilgiler en doğru şekilde kullanılarak konuşmanın üretilmesi sağlanır. Burada daha önce toplanmış ses veritabanından elde edilen parametrelerin doğru olarak seçilmesi ve bu parametrelere gerekli süre, tonlama ve genliklerin kazandırılıp en doğru şekilde bürünsel etkileri verip sentezleme işleminin gerçekleşmesi sağlanır. Metin ve dil analizi Doğal Dil İşleme (DDİ) ile gerçekleştirilmektedir. DDİ modülü ile en basitinden metnin fonetik transkripsiyonu elde edilmeli ve vurgu, süre, tonlama gibi sentezlemede kullanılacak parametreler elde edilmelidir. Bürün ve konuşma üretimi aşamasında ise Sinyal İşleme tekniklerinden istifade edilmektedir. Sentezleme Modülü olarak ta adlandırılabilen bu modül sayesinde de yukarıda da bahsedildiği üzere fonetik ve bürünsel bilgiyi kullanarak konuşmayı sentezler. Bu bilgileri kullanarak bir MKS sistemini aşağıdaki şekilde genişletmemiz mümkündür.

- 1) Ses Veritabanının Hazırlanması:
- 2) Doğal Dil İşleme Modülü:
  - a) Metin Analizi: Metnin yapısı belirlenir ve normalleştirme yapılır.
  - b) Fonetik Analiz: Normalleştirilmiş metnin fonetik açılımı belirlenir.
  - c) Bürünsel Analiz: Bürünsel özelliklerin uygulanacağı kısımlar ve parametreler belirlenir.
- 3) Sentezleme Modülü:
  - a) Bürün Üretimi: Süre bilgisi, tonlama bilgisi, perde frekans bilgisi gibi konuşmanın kalitesini belirleyecek bilgiler fonetik bilgiye eklenir.
  - b) Konuşma Sentezleyici: Konuşmanın kural tabanlı, eklemeli veya başka bir yöntem sayesinde sentezlenmesini sağlar [51].

MKS sisteminin üç temel birimini içeren ayrıntılı bir sistem mimarisi Şekil 5.2'de verilmiştir. Bu birimlerle ilgili bilgiler Bölüm 5.3'te verilmiştir.



Şekil 5.2. MKS Sisteminin Genel Mimarisi [51]

### 5.3. Bütüncenin Analizi ve Ses Veritabanının Oluşturulması

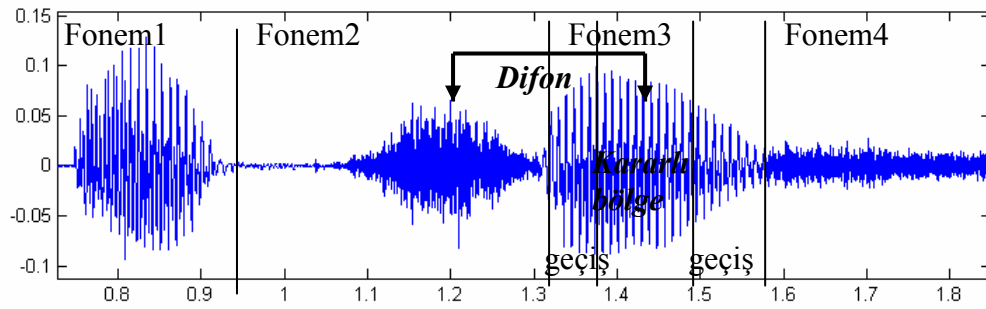
#### 5.3.1. Veritabanı tasarımı

Sistem tasarımında üzerinde durulması gereken ilk önemli nokta konuşma veritabanının nasıl oluşturulacağına karar verilmesidir. Sentez işlemi, veritabanı içerisinden alınacak işaret parçalarının işaret işleme algoritmalarıyla işlenerek bürünsel özelliklerinin değiştirilmesi ve ardarda eklenmeleri sonucu gerçekleştirileceği için veritabanı kalitesiyle sentezlenen konuşma işareti kalitesi arasında doğrudan bir ilişki vardır. Buna ek olarak diğer modüllerin (özellikle Parça Seçici modülünün) tasarımında da veritabanı içeriğini göz önünde bulundurmak gerekli olmaktadır.

Veritabanı tasarımında karar verilmesi gereken ilk nokta kullanılacak en küçük parçanın (ses biriminin) ne olması gerektiği ve veritabanının kapsayacağı ses uzayıdır. Bu kararın sağlıklı bir şekilde verilebilmesi için parçaların ardarda diziliminde oluşacak zorlukların göz önünde bulundurulması gerekmektedir.

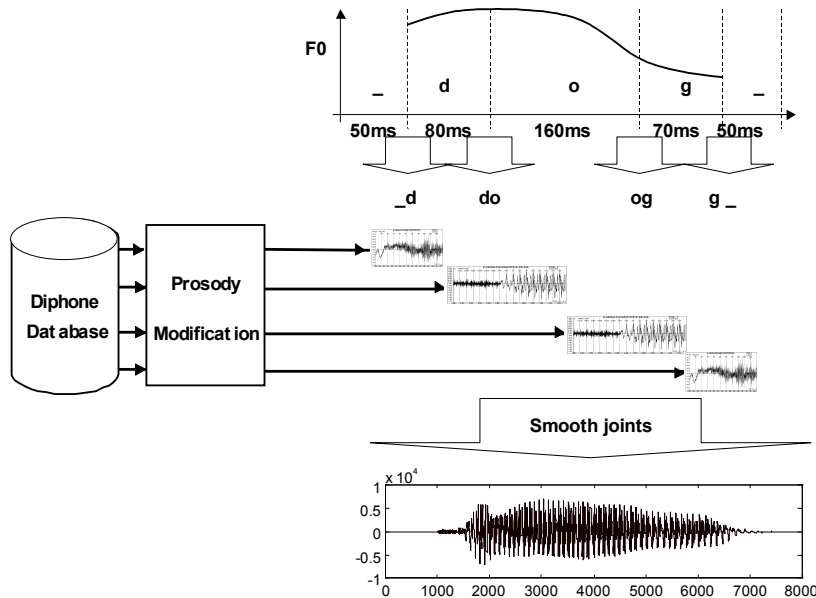
Şekil 5.3'te bir konuşma işareti görülmektedir. Konuşma işareti içeriğindeki fonemler öbekler halinde rahatlıkla gözlenebilmektedir (dört adet fonemin sınırları şekil üzerinde gösterilmiştir). Her bir fonem, ortada kararlı bir bölge, kenarlarda da geçiş bölgeleri içermektedir (Fonem 3 için bu bölgeler işaretlenmiştir).

Ses parçalarının ardarda dizilmesiyle ilgili yapılan çalışmalar göstermiştir ki geçiş bölgelerinin dinamik yapısı nedeniyle bu bölgelerde birleştirme işlemleri sırasında işitilebilir düzensizlikler oluşabilmektedir. Bu sebeple geliştirilen bir çok sistem, seslerin kararlı bölgelerinden kesip birleştirilmesi prensibine dayanmaktadır. Bu bölgelerde işaret işleme algoritmaları etkili bir şekilde kullanılıp kaynaştırmalar daha yüksek kaliteyle yapılabilmektedir.



Şekil 5.3. Konuşma işareti üzerinde ikili sesin işaretlenmesi

Konuşma işaretleri kararlı bölgelerin ortalarından parçalandıklarında ikili sesler elde edilmiş olur (Şekil 5.3 üzerinde bir ikili sesin sınırları oklarla gösterilmiştir.). Bugün ticari ve akademik olarak oluşturulmuş konuşma sentezcilerinin büyük kısmı Şekil 5.4’de gösterilen ikili seslerin ardarda dizilmesi prensibine dayanmaktadır. Veritabanı oluştururken ikili sesin temel ses birimi olarak seçilmesi yaygın bir uygulamadır. Bizim sistemimizde de aynı ses biriminin kullanılmasına karar verilmiştir. Bu kararı aldıktan sonra yapılması gereken, kaydedilmesi gereken ikili ses uzayının belirlenmesidir.



Şekil 5.4. İkili sesler kaynaştırılarak sentezleme yapılması

Sisteme iletilecek herhangi bir metnin sentezlenebilmesi için gereken ikili ses sayısı rahatlıkla fonem sayısından(n),  $n*n$  olarak hesaplanabilir. Anlamli veya anlamsız bütün metinlerin sentezlenmesi  $n*n$  ses ikilisi ile gerçekleştirilebilir. Çoğu dil için fonem sayısı 30-40 aralığında olduğu için ikili sesteki sentez yapan sistemler için 1500 civarında ikili ses kaydetmek yeterli olmaktadır.

Bu çalışmada da ikili ses tabanlı bir sistem kullanılması öngörülmüş ve bu amaçla bir bütüncü belirlenmiştir. Basit-çalışır sistemimizin veritabanının tüm ikili seslerin kapsamı hedeflenmiştir. Fonem sayısı 30 (29 + sessizlik) alınmış, bunun karşılığında 900 (30x30) adet ikili ses içerecek bir liste oluşturulmuştur. Bu şekilde tasarlanacak bir veritabanının yeterli kaliteyi sağlayamayacağı düşünülerek sentezin daha başarılı olması için Bütüncenin Türkçede en sık kullanılan üçlü seslerden oluşması kararlaştırılmış, bu üçlü seslerin içinde yer almayan ikili seslerin de bu listeye eklenmesi kararlaştırılmıştır. Teorik olarak 27000 (30x30x30) adet üçlü ses olmasına rağmen yapılan istatistiksel çalışmalar neticesinde 22327 adet farklı üçlü ses olduğu görülmüştür. Bu çalışmalar edebi ve teknik kitaplardan toplanan yaklaşık 5,15 M kelimedenden oluşan bir bütüncü üzerinde gerçekleştirilmiştir. Bu bütüncüde toplam 326620 adet tekil kelime olduğu görülmüştür. Üçlü ses istatistikleri bu tekil kelimeler üzerinden yapılmış ve en sık geçen 3000 adet üçlü ses belirlenmiş, bu üçlü seslerin Türkçe'yi %90 oranında temsil ettiği görülmüştür. Bu üçlü seslerde bulunmayan 383 adet ses ikilisi de listeye eklenerek toplam 3383 adet parçadan oluşan bir bütüncü oluşturulmuştur. Bütüncü ile ilgili istatistikler Tablo 5.1, ilk 20 adet üçlü ses bilgisi de Tablo 5.2'de verilmiştir. Üçlü seslerde sessizlik fonemi “#” simgesiyle gösterilmiş ve üçlü seslerin kelime başında, ortasında ve sonunda olma durumları bu şekilde birbirinden ayrılmıştır.



Tablo 5.1. Bütünce istatistikleri

Açıklama	Değer
Toplam Kelime Sayısı	5151572
Tekil Kelime Sayısı	326620
Toplam Cümle Sayısı	577454
Toplam üçlü ses Sayısı	13733843
Tekil üçlü ses Sayısı	22327
1 defa geçen üçlü ses Sayısı	2872
10 defadan fazla geçen üçlü ses sayısı	13544
10 defadan fazla geçen üçlü seslerin oranı	% 99,8
En fazla geçen 3000 üçlü sesin oranı	%90,6

Tablo 5.2. En fazla geçen üçlü sesler

No	Üçlü ses	Sayı	Oran	Toplam Oran
1	lar	129685	0,009442732	0,009442732
2	ler	118507	0,00862883	0,018071562
3	eri	81907	0,005963881	0,024035443
4	arı	77120	0,005615326	0,029650769
5	in#	72256	0,005261164	0,034911933
6	ini	62483	0,004549564	0,039461497
7	ın#	57516	0,004187903	0,0436494
8	an#	56337	0,004102057	0,047751456
9	en#	55624	0,004050141	0,051801597
10	#ka	53976	0,003930145	0,055731742
11	ile	50148	0,003651418	0,05938316
12	ama	49832	0,003628409	0,063011569
13	ını	49500	0,003604235	0,066615804
14	rin	47662	0,003470405	0,07008621
15	bil	45978	0,003347788	0,073433998
16	rn	42270	0,003077798	0,076511796
17	yor	41899	0,003050785	0,079562581
18	lan	41633	0,003031417	0,082593998
19	ala	41032	0,002987656	0,085581654
20	nda	39769	0,002895694	0,088477348

### 5.3.2. Veritabanının kaydedilmesi:

Hazırlanan bu bütüncede ikili ve üçlü seslerin kaydedilmesi esnasında söyleniş zorluklarının yaşanmaması ve kararlı kayıtların alınabilmesi için bu listeden logotomlar oluşturulmuştur. Bu amaçla hazırlanan bir liste TABLO 5.3 de verilmiştir. Logotom kurallı fakat anlamsız kelime demektir. Bu kelimelerin oluşturulması için ikili ve üçlü seslerin başına ve sonuna dolgu sesler yerleştirilmiştir. Bu dolgu sesler yerleştirilirken etiketleme esnasında sorun yaşanmaması için /k/,/p/ ve /t/ gibi ötümsüz seslerin kullanılmasına ve ünlü uyumuna dikkat edilmiştir.

Tablo 5.3. En sık kullanılan üçlü sesler ve logotom listesi

No	Üçlü ses	Logotom
1	Lar	balarat
2	Ler	beleret
3	Eri	ederidi
4	Arı	adaradı
5	in#	idin#
6	İni	idinidi
7	ın#	ıdn#
8	an#	adan#
9	en#	eden#
10	#ka	#kada
11	İle	idilede
12	Ama	adamada
13	Inı	ıdnıdı
14	Rin	birinit
15	Bil	bibilit
16	Rın	bırınıt
17	Yor	boyorot
18	Lan	balanat
19	Ala	adalada
20	Nda	bandada

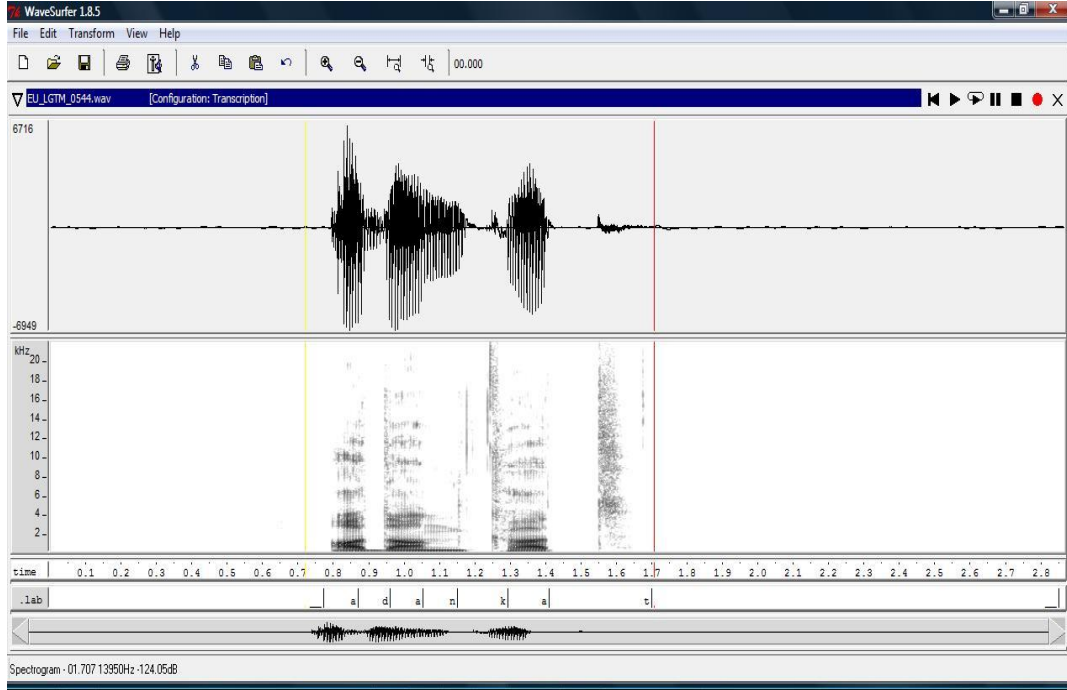
Hazırlanan logotomlar Şekil 5.5 de gösterilen TÜBİTAK – UEKAE yarı yansız odalarında kaydedilmiştir. Kaydetme formatı olarak 44.1 KHz ve 16 bit PCM, mono kodlama seçilmiştir. Kaydetme esnasında yanlış okumalar tekrarlanmış ve tüm kayıt esnasında aynı ses tonunun korunmasına dikkat edilmiştir.



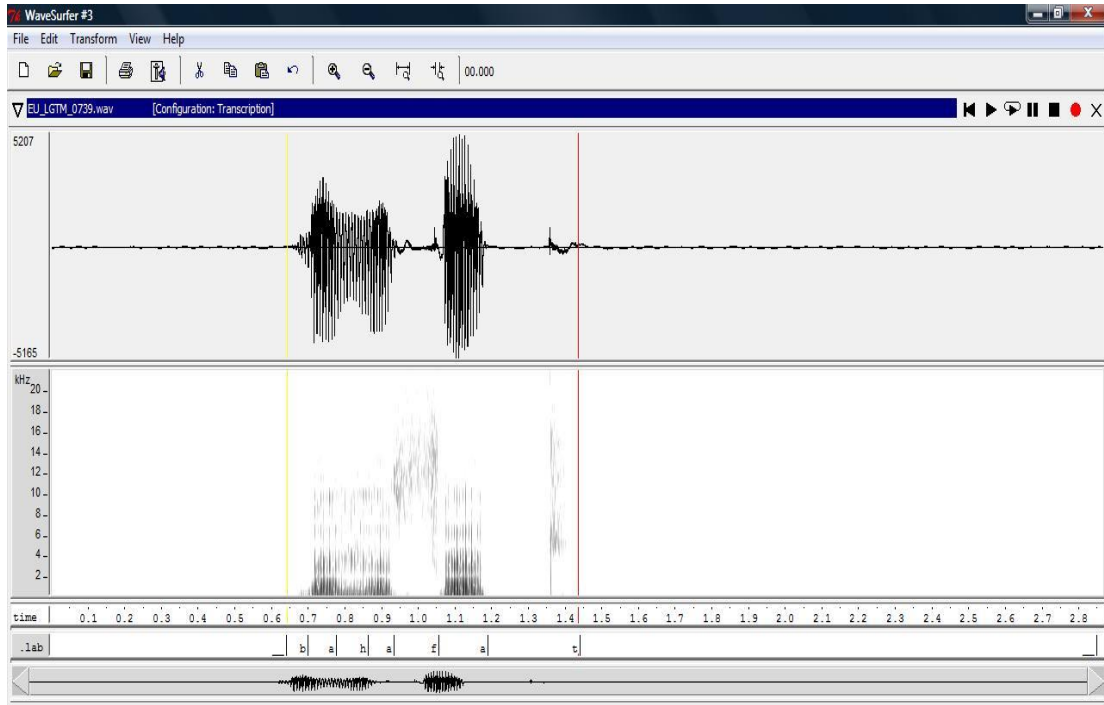
Şekil 5.5. Yarı Yansız Kayıt Odası

### 5.3.3. Veritabanının Etiketlenmesi

Kaydedilen veritabanının etiketlenme süreci sentezleme sistemlerinin en zorlu aşamalarından birini oluşturmaktadır. Bu çalışmada her bir kaydın elle sıfırdan etiketlenmesi yerine Sphinx konuşma tanıma sisteminden [15] istifade edilerek kayıtlar otomatik olarak etiketlenmiştir. Etiketlenen bu dosyalar daha sonra wavesurfer [52] programı yardımıyla elle kontrol edilerek hataları düzeltilmiştir. Örnek etiketlemeler Şekil 5.6 (a) ve (b) de görülebilir.



(a)



(b)

Şekil 5.6. Wavesurfer kullanılarak yapılan etiketlemeler

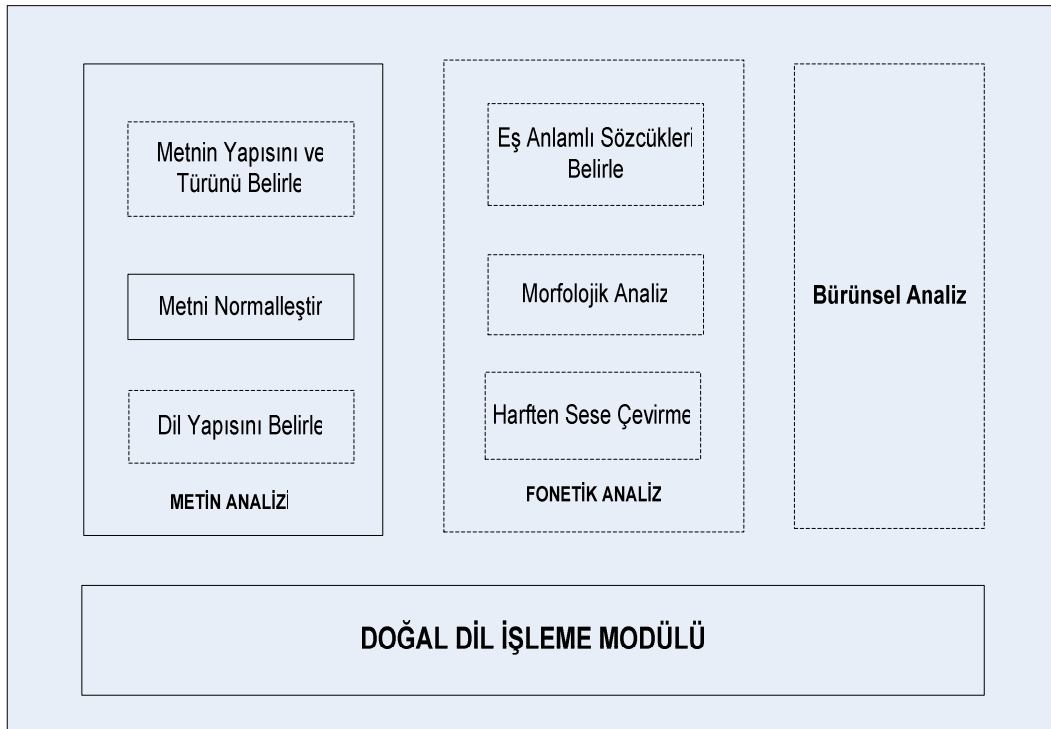
- a. **adankat** logotomunun etiketlenmesi
- b. **bahafat** logotomunun etiketlenmesi

### 5.3.4. Veritabanından, "Sentezleme modelinin" geliştirilmesi

Etiketlenen bütün veriler ilgili fonemlerin başlangıç ve bitiş sürelerini içerecek şekilde indekslenerek bir dosyaya kaydedilmiştir. Bu dosya model dosyası olarak kullanılacaktır. Hedef metinle ilgili ikili ve üçlü sesler bu dosyadaki bilgiler yardımıyla belirlenerek ilgili kısımlar kaynak dosyadan çekilecektir. Daha sonra bu parçalar birbirlerine eklenmek suretiyle sentezleme işlemi gerçekleştirilecektir.

### 5.4. Doğal Dil İşleme Modülü

Şekil 5.7 'de DDİ modülünün ayrıntısı verilmiştir. Bu çalışmada gerçekleştirilen basit -çalışır sistemde kesikli çizgilerle gösterilen kısımlar sisteme eklenmemiştir.



Şekil 5.7. Doğal Dil İşleme Modülü

### **5.4.1. Metin analizi**

Buradaki yapılacak işleri de bir kaç kısma ayırmamız mümkündür:

#### **5.4.1.1. Metnin yapısının tespit edilmesi**

Bu kısımda ilk önce metnin yapısı iyice analiz edilmelidir. Cümle başı, paragraf başı, noktalama işaretleri belirlenip yorumlanmalıdır. Bu aşama bürünel bilgilerin üretilmesi için gereklidir. Örneğin bir cümleye başlarken daha yüksek perdeden başlanacağı için bu bilginin saklanması gerekmektedir. Gene noktalama işaretlerine göre cümlenin bitişindeki vurguyu farklı şekillerde vermek mümkündür. En basitinden bir soru cümlesi veya bir ünlem cümlesi birbirinden farklı yorumlanmalıdır [51].

#### **5.4.1.2. Metin normalleştirme**

Rakamlar, kısaltmalar ve semboller bu kısımda tespit edilmeli ve bunların açılımı yapılmalıdır. Örneğin “%5” gibi bir yazımla karşılaşınca bu “Yüzde Beş” olarak çevrilmeli ve sonraki aşamalara bu şekilde geçirilmelidir [51].

#### **5.4.1.3. Dil yapısının belirlenmesi**

Bu kısımda da kelimelerin, tamlamaların, cümleciklerin ve cümlelerin semantik(anlamsal) ve sentaktik (sözdizimsel) analizi yapılmalıdır. Bu kısımda bir kelimenin hangi gramatik yapıda ve hangi görevde kullanıldığı belirlenir ve kullanıldığı yere ve göreve göre fonetik açılımının yapılması ve bürünel parametrelerinin saptanması sağlanır [51].

### 5.4.2. Fonetik analiz

Bu kısımda metnin yazılı sembollerden, vurguları da içine alacak şekilde konuşma sembollerine çevrilir. Bu kısmı da üç bölümde inceleyebiliriz:

#### 5.4.2.1. Eş anlamlı sözcüklerin tespiti

Buradaki asıl olay yazılışı aynı olmasına rağmen okunuşları farklı olan sözcüklerin tespit edilmesi problemdir. İsim, fiil veya farklı bir görev yüklenmiş bir kelime her görevde farklı vurgu ile okunabilir. Bu problemin kaldırılması için kelimenin hangi görevde kullanıldığı tespit edilip vurgu bilgisi ona göre verilmelidir. Bu aşama Türkçe için büyük bir problem teşkil etmektedir. N-gramlarla belli bir olasılığa göre bir sonuca gitmek mümkündür. Örnek olarak ‐Aydın‐ kelimesini alırsak bu kelimenin il ismi mi, özel bir isim mi yoksa ‐Gözün aydın‐ derken fiil olarak mı kullanıldığı tespit edilmeli ve vurgusu belirlenmelidir.

#### 5.4.2.2. Morfolojik analiz

Morfolojik analiz, kelimeyi eklerine ve köklerine ayırır. Bu sayede de kökün aldığı ekler analiz edilerek vurgunun hangi hece üzerinde olacağı tespit edilir. Tabii olarak basit bir heceleyici de içermek zorundadır [53].

#### 5.4.2.3. Harften sese dönüştürme

Fonetik analizin son kısmını oluşturur. Sözlükten bakma (look up) mantığı veya Harften Sese Dönüştürme (LTS "letter- to- sound") kuralları belirlenerek veya ikisi bir arada kullanılarak yapılır. Bu kurallar ilgilenilen dilin özelliklerine göre belirlenir. Mesela Türkçede ‘ğ’ harfi ünsüz yumuşamasıyla ortaya çıkmışsa okunmaz, ünlüyü kaydırır. ‐aldığı‐ ve ‐kaydığı‐ örneğinde görüldüğü gibi. Bunu gibi yüzlerce kural tanımlamak mümkündür.

Sözlük Tabanlı Çözüm: Mümkün olduğu kadar fazla veri depolanır. Sözlükte bulunmayan sözcükler belirlenmiş kurallarla fonetik olarak açılabilir. Türkçe için sonlu durum makinaları ile geliştirilmiş, tam fonetik çeviri yapan sistemler geliştirilmiştir.

Kural Tabanlı Çözüm: Fonetik kurallara göre fonetik açılım yapılır. Bu kural kapsamındaki kelimeler sözlüğe alınmaz. Her dil için en fazla kullanılan kelimeler tespit edilerek istatistiksel modeller de çıkarılabilir.

### 5.4.3. Bürünsel analiz

Metin ve fonetik analiz bilgilerinden istifade edilerek hedef metni oluşturacak parçalara verilecek bürünsel etkiler bu bölümde belirlenir. Hangi heceye vurgu verileceği, hangi sesin süresinin uzatılacağı, cümlenin tonlamasının nasıl olacağı (soru, biten veya süren ezgi) bu analiz neticesinde cevaplanır.

Şekil 5.7'de gösterilen üzere "Metin normalleştirme" ve "Harften Sese Dönüştürme" birimleri ile ilgili olarak yapılan detaylı çalışmalar Bölüm 5.5 ve 5.6 'da gösterilmiştir.

### 5.5. Metin Normalleştirme

Geliştirilen sentezleyiciler girdi olarak sadece alfabetik karakterleri kabul etmektedir. Bir dil için geliştirilen bir sentezleyici sadece o dilde kullanılan alfabedeki harfleri kabul edecektir. Türkçe için de bu alfabemizde bulunan 29 harfe karşılık gelmektedir.



Yazılı metinler incelendiğinde kısaltmalar, rakamlar, semboller, yabancı kelimelerin kendi orijinal dillerindeki gibi yazıldığı durumlar olduğu görülecektir. Tüm bu kelimelere kısaca “Standart Olmayan Kelimeler” denilmektedir ve bu kelimeler sentezleyicinin kabul edeceği harflerle ifade edilmelidir. Bu işleme kısaca metin normalleştirme denilmektedir.

Metinden konuşma sentezleme sistemleri, sentezleme için uygun sesleri seçerken, belirlenen seslere göre genişletilmiş metinleri kullanmaktadır. Bir metindeki mevcut bütün bilgilerin bu sesler cinsinden ifade edilmesi gerekmektedir. Metnin bu seslere açılması için harften sese dönüşüm yöntemlerinin uygulanması gerekmektedir. Bu yöntemlerin uygulanabilmesi için metnin içinde harf harici herhangi bir bilginin olmaması gerekir. Mevcut metinler incelendiğinde ise rakamlar, kısaltmalar, semboller, sayılar, numaralar gibi standart olmayan kelimelerin metinde olduğu görülür. Standart olmayan bu kelimelerin tam açılımlarının yapılarak harflerle ifade edilmesi için metnin normalleştirilmesi gerekir. Örneğin “1978” sayısı “bin dokuz yüz yetmiş sekiz” olarak, “190 cm.” ise “yüz doksan santimetre” olarak açılmalıdır.

Metin normalleştiricinin görevi metindeki standart olmayan kelimeleri tespit ederek bunların harf bazında açılımlarını yapmaktır. Harften sese dönüşüm yapılmadan önce metnin harf bazında otomatik olarak açılması gerekmektedir. “Metin Normalleştirme Modülü” olarak adlandırılan bu modül kural tabanlıdır. Sistemin ana mantığı verilen metindeki standart olmayan kelimeleri tespit ederek bu kelimeleri belirlenen kurallara ve listelere göre açmaktır.

Bir metin normalleştirme işleminde önce metni jetonlarına ayrılır, sonra da bu jetonların her birinin standart olmayan kelime olup olmadığına bakılır, eğer normalleştirilmesi edilmesi gereken bir jeton ise hangi kurallar göre normalize edileceğini belirlenir ve metin bu kurallara göre alfabetik karakterlere çevrilerek normalleştirilir.

Türkçe yazım kuralları incelenerek “Standart olmayan kelimeler” belirlenmiştir ve Tablo 5.4’te sunulmuştur.

Tablo 5.4. Standart olmayan kelimeler ve örnekler

Standart Olmayan Kelime	Türkçe Örnek
Kısaltmalar	km, ist.,cad.
Harf dizisi	UEKAE, TBMM
Doğrudan okunan kısaltmalar	TÜBİTAK
Yazım Hataları	harta (harita)
Sayılar –esas sayılar-	13, 46, 0.1, 2/2
Sıra bildiren sayılar	3., I. MURAT, II. OSMAN, 29 may.
Telefon Numaraları	262 648 14 30
Rakam bildiren sayılar	Oda 304
Belirteç, tanımlayıcı	MILSEC2, 3M, ML380
Adreslerde kullanılan sayılar	Küçükyalı 63
Posta kutusu ve posta kodu	PK.74 41470
Zaman belirteci	20:45, 14:30
Tarih Belirteci	08/08/2008, 7/7/07
Yıl bildiren sayılar	80ler, 1995, 2000ler
Para miktarı bildiren sayılar	230 YTL, 345\$
Çok yüksek para miktarları	230 milyon TL
Oranlar	%13, %5.71
Söyleyişi olmayan Kelimeler , kelime sınırları	***bkz,
Söyleyişi olmayan, cümlecik sınırları	geçmiş zaman olur ki..
Neşeli yazımlar	akşamdaaaaaaaaaaaaaaannn akşama
url, path adresi veya email	www.uekae.tubitak.gov.tr, /usr/tts, bicil@uekae.tubitak.gov.tr
Dikkate alınmayacaklar	Ascii ile yapılmış şekiller, yanlış format seçilmesinden dolayı bozulan karakterler

Normalleştirme modülü kuralların tanımlanabileceği ve kısaltma listelerinin sonradan eklenebileceği, devamlı geliştirilebilecek bir modül olarak gerçekleşmiştir. Sisteme verilen metin cümlelere ayrılmakta, “Kurallı İfadeler (Regular expression)” ile yeni kurallar tanımlanabilmektedir. Şu an için esas sayılar, rakam bildiren sayılar,

sıra bildiren sayılar, kısaltmalar için gerekli kurallar belirlenerek sisteme eklenmiştir. Kısaltma listesi oluşturulurken TDK'nin listesi kullanılmıştır [54].

### 5.6. Harften - Sese Dönüştürme Kuralları

Her dilde yazılı dilin kurallı yapısı karşısında konuşma dili, geniş bir çeşitlilik gösterir. Bu nedenle, toplumsal uzlaşmayı yansıtan ve gerekli durumlarda söyleyişte birleştirici olma niteliğiyle belirginleşen ölçünlü söyleyiş kurallarının saptanması, dil kullanıcıları açısından açık değerler taşıyacaktır[39]. Metinden konuşma sentezleme sistemlerinde kullanılan sözlüklerin, bir dildeki bütün kelimeleri kapsamı çok fazla mümkün değildir [55]. Bu yüzden sentezleme sistemlerinde kelimelerin okunuşunu tahmin edecek yöntemlerin geliştirilmesi gerekmektedir.

Metinden Konuşma Sentezleme sistemlerinde eldeki yazı dili işaretlerinin konuşma dili işaretlerine çevrilmesi gerekmektedir. Bu amaçla look-up mantığıyla çalışan sözlük tabanlı bir yöntem kullanılacak olsa bile, burada ne kadar geniş kapsamlı bir sözlük kullanılırsa kullanılsın, yine sözlükte bulunmayan kelimelerle karşılaşılması muhakkaktır. Bu yüzden yazılı dilden konuşma diline geçecek kuralların belirlenmesi; hem sözlük harici karşılaşılabilecek kelimelerin, konuşma dili işaretlerine çevrilmesini sağlayacak, hem de kesin doğrulukla konuşma diline çevrilen kelimelerin sözlükten çıkarılarak sözlüğün hacminin küçültülmesi sağlanacaktır.

Çoğu dilde ortografik sistemle söyleyişler arasında dilin özelliğine göre ilişkiler mevcuttur. Bu ilişkiler bazı dillerde çok basit (İspanyolca gibi), bazılarında zor (İngilizce gibi) ve bazılarında da oldukça karmaşıktır (Japonca gibi) [55]. Türkçe yazı dili ile konuşma dili büyük ölçüde örtüşen bir dil olmakla beraber yine de her ortografik sembollerin kullanıldığı her yerde aynı fonetik bilginin taşınmadığı aşıkardır. Osmanlıca'dan miras kalan ve diğer dillerden hayatımıza geçmiş kelimeler Türkçe'nin fonetik yapısı içindeki istisnaları oluşturmakla beraber dilimiz kendi içinde söyleyiş kurallarını barındırmaktadır. Örneğin /k/, /l/ ve /g/ harfleri ön damak ünlüleri ile farklı, arka damak ünlüleri ile farklı söyleyişe sahip olmalarına rağmen,

kurallar çok kesin olduğundan alfabemizde tek ortografik sembolle (harfle) ifade edilmişlerdir. Söyleyiş ile ilgili daha ayrıntılı analizler için Bölüm 4.5'e bakılabilir.

Her ne kadar Sentezleme sisteminde böyle bir modülün bulunması sistemi iyileştirecek olsa da bizim geliştireceğimiz basit- çalışır sistemde her harfe karşılık bir ses temsil edildiğinden bu şekilde kurallardan oluşan bir modüle ihtiyaç duyulmamıştır. Ama gene de Türkçe için yapılan çalışmalar [56] ve sonuçları aşağıda anlatılmıştır.

### 5.6.1. Yöntem

Bu kuralların belirlenmesi için temel iki yöntem mevcuttur: Kural tabanlı yöntemler ve istatistiksel yöntemler. Her ne kadar kuralların elle belirlenmesi mümkün olsa da, sentezleme için girdi olarak kullanılacak kelimelerin her zaman Türkçe kökenli olmaması ve bu şekildeki kelimelerin dilimizdeki mevcudiyetinin çokluğu istisnai durumların sayısını çoğaltmaktadır. Bu yüzden tamamiyle istatistiksel bir yöntem tercih edilmiştir.

Bu istatistiksel çalışmada, Türkçe'nin Söyleyiş Sözlüğü'ndeki [39] kelimeler ve bunların fonetik açılımları esas alınmış ve Edinburgh Üniversitesi'nde geliştirilen EST [57] bünyesindeki araçlardan istifade edilerek Türkçe için kurallar belirlenmeye çalışılmıştır. Bu nedenle Türkçe için bir fonem kümesi belirlenmiş (bkz. Bölüm 4.5) , sözlükteki bazı ikilemeler ve yabancı kelimeler çıkarılmış, sözlük eğitim ve test olmak üzere (9:1 oranında) iki kısma ayrılmış, vurgu bilgileri kaldırılmış, fonetik açılımlarda kullanılan uzatma işaretleri yerine ünlü harflerin kısa ve uzun hallerini ayrı ayrı belirleyen fonemler kullanılmıştır.

Alfabedeki her harf için ayrı bir CART ağacı eğitilmiştir. Bu eğitim esnasında her harfin sağındaki ve solundaki dört harf dikkate alınarak üretilen bir öznitelik vektör kümesi kullanılmıştır. Ağaçların eğitilebilmesi için kullanılacak öznitelik vektörlerinden bir örnek Tablo 5.5'te verilmiştir. Tablodaki “#” işareti kelime başlangıç ve bitiş noktalarını göstermektedir.

Tablo 5.5. Maydanoz kelimesi için elde edilen öznitelik vektörü

M	0 0 0 # m a y d a nil
a	0 0 # m a y d a n nil
y	0 # m a y d a n o nil
d	# m a y d a n o z nil
a	m a y d a n o z # nil
n	a y d a n o z # 0 nil
o	y d a n o z # 0 0 nil
z	d a n o z # 0 0 0 nil

CART ağacının oluşturulması sırasında, her harfin sağında ve solunda mevcut harflere göre sorular sorulmuş ve cevapların öznitelik vektör kümesinde istenilen sayıda bulunduğu durumlardaki sorular ağaca alınmıştır. Burada önemli olan husus “durma değeri” olarak adlandırabileceğimiz, sorunun cevabının eğitim kümesinde bulunma sayısıdır. Durma değeri düşük olarak seçildiğinde, ister istemez modelimizin büyüklüğü artacaktır. Modelimizin büyüklüğü CART ağacındaki soru sayısı ve yaprakların sayısının toplamıdır [55].

### 5.6.2. Deneyler

Eğitim kümesi toplam 27072 kelimedenden, test kümesi de 3009 kelimedenden oluşmaktadır. Test kümesi ,Türkçenin Söyleyiş Sözlüğündeki her 10 kelimedenden biri ayrılarak oluşturulmuştur. Herhangi bir kuralın ağaca girebilmesi için test kümesinde görülme sayısı, yani durma değeri 9 olarak verilmiştir. Soruların derinliği belirlenirken harfin önündeki ve arkasındaki 4 harf dikkate alınmıştır. Bu parametrelere göre elde edilen sonuçlar Tablo 5.6’da , harf bazındaki istatistiklerde Tablo 5.7’te verilmiştir.

Tablo 5.6. Türkçe söyleyiş sözlüğü için elde edilen sonuçlar

Durma Değeri	Kelime	Harf	Ağaç Ebatı
9	85.04	97.59	1331

Tablo 5.7. Harf bazlı sonuçlar

Harf	Toplam sayı	Doğru sayısı	Oran (%)	Harf	Toplam sayı	Doğru sayısı	Oran (%)
A	2838	2632	92,74	M	1229	1229	100
B	464	464	100	N	1050	1031	98,19
C	308	308	100	O	634	627	98,90
Ç	322	322	100	Ö	185	181	97,84
D	495	494	99,80	P	331	331	100
E	1931	1842	95,39	R	1287	1286	99,92
F	217	217	100	S	826	826	100
G	292	286	97,95	Ş	404	404	100
Ğ	169	169	100	T	1070	1070	100
H	314	310	98,73	U	612	584	95,42
I	890	888	99,78	Ü	445	441	99,10
İ	1609	1542	95,84	V	241	235	97,51
J	45	45	100	Y	513	508	99,03
K	1773	1735	97,86	Z	434	433	99,77
L	1525	1471	96,46				

Kurallar incelendiğinde ünsüz harflerin ünlü harflere göre daha iyi modellendiği görülecektir. Hatta b,c,ç,d,f,j,m,p,s,ş,t ünsüz harflerinin tek bir kuralla modellendiği görülmüştür.

Karar ağacı modeli kullanılarak söyleyiş kurallarının belirlenmesiyle görülmüştür ki Türkçe'nin kurallı bir dil olması nedeniyle küçük boyutlardaki ağaçlarla bile iyi modelleme yapmak mümkündür. Bu model ile oluşturulacak kurallar kapsamına giren sözcüklerin “Metinden Ses Sentezleme” sistemlerinde kullanılacak sözlüklerden çıkarılarak sözlük hacmi ciddi oranda düşürülebilir.

Özellikle ünlü harflerin doğruluk oranının düşük olması, hem bu harflerin karşıladığı fonem sayısının fazla olmasından, hem de /ğ/ ile beraber kullanıldığı durumda bu seslerin uzun okunacak olmasından dolayı kuralların karmaşıklaşmasının bir sonucu olarak görülmektedir. Ayrıca sözlük bünyesinde bulunan yabancı kaynaklı kelimeler de bu oranın düşmesine sebebiyet vermektedir. Kurallar belirlenirken

kelimelerin kökeni ve görevi göz önüne alınmamıştır. Bu hususların da eklenmesinin daha sağlıklı kurallar oluşturacağı açıktır.

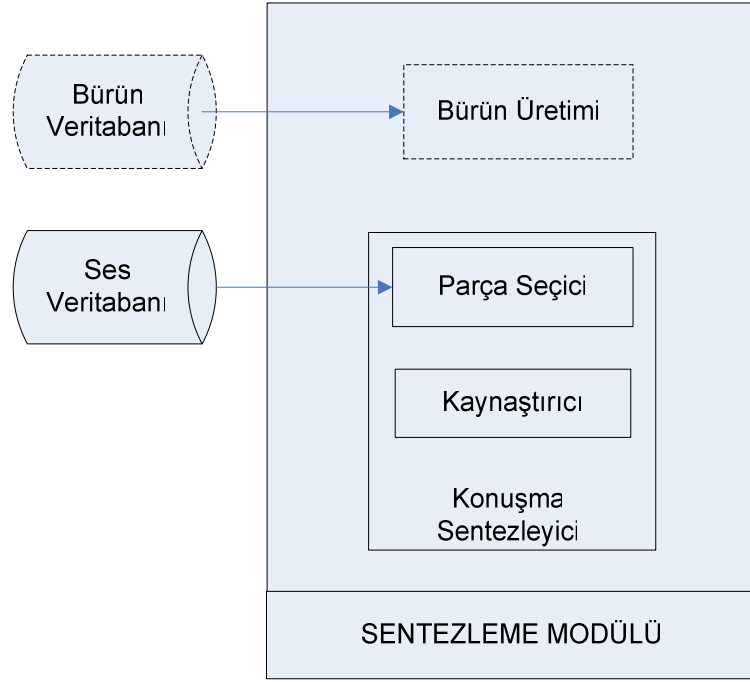
Sonuç olarak, bu kurallar ile MKS'de kullanılan sözlük hacmi küçültülerek özellikle bellek büyüklüğü ve işlemci gücü açısından zayıf mobil cihazlar üzerinde geliştirilecek uygulamalarda avantaj sağlanabilir. Bununla birlikte, deneysel sonuçlar göstermiştir ki, Türkçe'de düzensiz durumların ve harflerin karşıladığı fonem sayısının çok fazla olmaması ve konuşma dili ile yazı dili arasında çok fark bulunmaması nedeniyle Türkçe başarı oranları diğer dillere göre daha yüksektir.

Bu tez kapsamında gerçekleştirilecek sentezleme sisteminde harf ses ayırımına dikkat edilmediğinden söyleyiş kuralları sisteme entegre edilmemiştir.

### **5.7. Sentezleme Modülü**

Sentezleme Modülüne ilişkin ayrıntılar Şekil 5.8'da gösterilmiştir.

Gerçekleştirilen basit- çalışır sistemde her ne kadar bürünsel bir sentezleyici düşünülerek tasarım yapılsa da, bürün üretimi ile ilgili kısım sisteme eklenmemiştir. Bu yüzden bürünle ilgili kısımlar kesikli çizgilerle gösterilmiştir. Bürünsel etkiler konuşmanın kalitesini birebir etkileyecek özelliklerdir. Türkçe için ayrıntılı bürünsel analiz çalışmaları henüz yeterli düzeyde yapılmamıştır. Bürünsel etkilerin nasıl tanımlanacağı hususunda diğer dillerde "Tone and Break Indices" (TOBI) [58] gibi çalışmalar yapılmış ve TOBI dikkate alınarak etiketlemeler yapılmış olmasına rağmen Türkçe'nin henüz bürünsel bir veritabanı oluşturulmamıştır. Bu konuda Oskay'ın yaptığı çalışma incelenmiş ve TOBI indislerinden sadece iki tanesinin (H ve L) kullanıldığı görülmüştür [59]. Bürün analizi ve üretimi konuları gelecek çalışmalar içinde yapılmasına karar verilmiştir.



Şekil 5.8. Sentezleme Modülü

### 5.7.1. Bürün üretimi

Bürün (prozodi) genel olarak vurgu, tonlama, süre gibi özelliklerin tamamına verilen genel isimdir [60]. Kısacası konuşmanın doğallığı bürünsel analizinin ne kadar doğru yapıldığı ve bu analiz neticesinde ne derece doğru etkilerin verildiği ile sıkı bir bağlantı içindedir. Yine her dilin ses özelliklerine göre bürünsel etkileri. Konuşma sentezlemede bürün üretimi için özel fonksiyonlar oluşturulmalıdır. Bürünsel özelliklerden en bileşenlerinden biri vurgu olup, kelimenin anlamını vurgunun verildiği heceye göre değiştirmek mümkündür. Örneğin “ağrı” kelimesinde vurgu ilk hecede verilirse şehir ismi, son hecede verilirse acı manası verilmiş olacaktır. Bu ayırt edici özellik dilbiliminde supra-segmental özellik diye geçmektedir [33].

Bürünsel özellik vurgunun yanı sıra konuşma sinyallerindeki perde frekansı, genlik ve hece uzunluğu ile de ilgilenir [60]. Bunlar da konuşmadaki duyguları açıklamayı sağlar.



### 5.7.2. Konuşma sentezleyici

İnsanda konuşma üretimi, akciğerlerden pompalanan havanın ses telleri, dil, diş, damak, dudak yardımıyla şekillendirilip ağız veya burundan çıkması neticesinde gerçekleşen kompleks bir süreçtir. Her fonemin oluşması için farklı kombinasyonların gerçekleşmesi ve fonemlerin bir araya gelip hecelerin oluşturması ve hecelerin kelimeleri, kelimelerin de cümlelerin oluşturması peş peşe gerçekleştirilen bu kombinasyonlarla mümkündür. Bilgisayarda konuşma sentezlenirken de konuşma üretilirken bu sistem göz önüne alınarak bir analogi geliştirilmiştir. Her bir etken farklı parametrelerle sisteme entegre edilerek konuşma sentezleme metotları geliştirilmiştir. Bunun için önerilen iki yöntem mevcuttur:

#### 5.7.2.1. Kural tabanlı / formant sentezleyici yöntemi

Kural tabanlı sistemlerin iki ana bileşeni mevcuttur. Bunlardan biri sinyal diğeri de konuşma yolunu simüle eden bir filtredir. Yani öncelikle bir sinyal üretilir, daha sonra bu sinyal dinamik bir filtreden geçirilerek konuşma üretilir. Filtrenin parametreleri tamamıyla akustik özellikler yardımıyla belirlenir. Burada ünlü ve ünsüz fonemlerin üretilmesi akustik açıdan farklı olduğu için sentez aşamasında da farklı özellikler içerecektir. En basitinden ünlü bir fonemler üretilirken periyodik bir sinyale ihtiyaç varken, ünsüz bir fonemin üretilmesi için rastgele bir gürültü sinyaline ihtiyaç bulunmaktadır. Konuşma sinyalleri sabit olmadığı için konuşma sentezleme süreci içerisinde frekans ve sinyal kaynağının süreç içerisinde değiştirilmesi gerekmektedir. Kural tabanlı sistemlerde konuşmanın perdesini, frekansını veya başka bir parametresini değiştirmek için kurallar kümesi tanımlanması önerilmektedir. Bu kurallar tamamen insanda konuşma üretim mekanizması göz önüne alınarak belirlenmelidir.

#### 5.7.2.2. Eklemeli sentezleme yöntemi

Bu yöntemde önceden kaydedilmiş konuşma parçacıklarının eklenerek konuşmanın sentezlenmesi söz konusudur. Burada da hangi konuşma parçacığının seçileceği ve

kaydedileceği büyük önem taşımaktadır çünkü sentezlenen konuşmanın kalitesi ve istenilen etkilerin verilmesi kullanılan konuşma parçacıkları ile sınırlıdır. konuşma parçacığının kalitesi ve sayısı arasında optimizasyon yapılmalı ve sistem ona göre tasarlanmalıdır.

Konuşma sentezleyici içindeki parça seçici ve kaynaştırıcı birimleri Bölüm 5.7.3'te açıklanmıştır.

### 5.7.3. Uygulanan sentezleme yöntemi

Bu çalışmada uygulanan yöntem birim seçme sentezi ile ikili ses sentezi arasında bir yöntemdir. Veritabanı tasarım kısmında anlatıldığı üzere üçlü sesler ve ikili seslerin birleşiminden istifade eden eklemeli bir yöntem geliştirilmiştir. Konuşma parçacığının birim uzunluğuna göre hacminin ve kalitesinin değişimi Tablo 5.8'de gösterilmiştir.

Tablo 5.8. Konuşma parçacıklarına göre hacim ve kalitesi

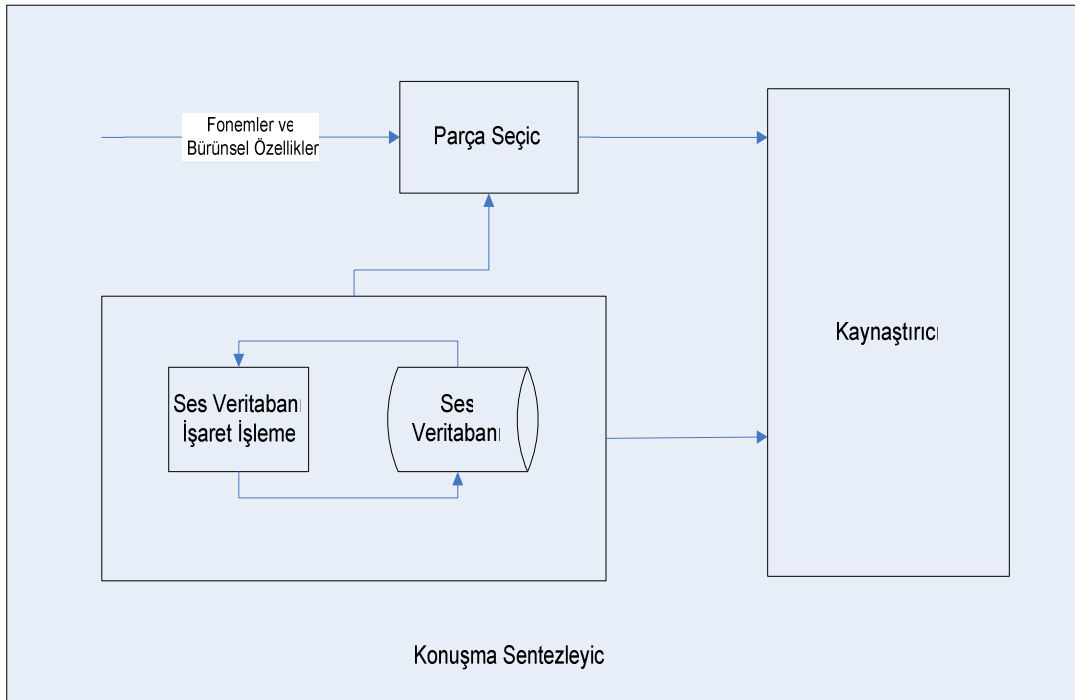
Birim Uzunluğu	Birim türü	Birim sayısı	Kalite
Çok Kısa	Fonem	44	Çok Düşük
Kısa	İkili ses	900	Düşük
Orta	Üçlü ses	27000	Orta
Uzun	Kelime	>100K	İyi
Çok Uzun	Cümle	$\infty$	Yüksek

Bu tablodaki fonem sayısı İclâl Ergenç'in [19] belirlediği fonemler esas alınarak oluşturulmuştur. Normalde alfabemizde 29 harf olsa da konuşma sırasındaki fonem sayımız daha fazladır. Tablodan da görüleceği üzere kalitenin artması için kullanılan birim sayısının da doğru orantılı olarak artması gerekmektedir. Fakat gerek depolama gerekse işlem hacmi sıkıntısından belli bir seçim yapılması gerekmektedir. Birim olarak fonem kullanılması, fonemlerin ardarda eklenerek konuşmanın sentezlenmesini öngörmektedir [61]. Fakat bunun neticesinde çok kötü bir sentez

karşımıza çıkmaktadır. Geliştirilen sistemlerde bu birimlerden başka ikili ses, üçlü ses, yarım-hece, hece, tamlama gibi birimler de önerilmektedir.

Bu tez kapsamında yapılan çalışmalarda Türkçe için iyi bir sentezleyicinin geliştirilmesi için asgari 44 adet fonemin kullanılması gerektiği öngörülmüştür. Ancak bu fonemlerle oluşturulacak veritabanının kaydedilmesi için profesyonel ses sanatçılara ihtiyaç duyulacağından 29 adet harf baz alınarak bir veritabanı belirlenmiştir. Hazırlanan veritabanı Bölüm 5.3'te açıklandığı üzere üçlü seslerden ve ikili seslerden oluşturulmuştur.

Şekil 5.9'da gerçekleştirilen sentezleyicinin akış şeması verilmiştir.



Şekil 5.9. Konuşma sentezleyici akış şeması

Konuşma sentezleyici ses veritabanından verilerin parça seçici yardımıyla seçildikten sonra birbirine eklemek suretiyle konuşmanın üretildiği kısımdır. Birim seçme, kaynaştırma gibi işlemler bu modül tarafından gerçekleştirilir. Bürün Üretimi ile beraber kullanıldığında bürünsel bir sentez elde etmek mümkündür. Ama bu haliyle sentezleyici sadece monoton konuşma sentezleyecektir.

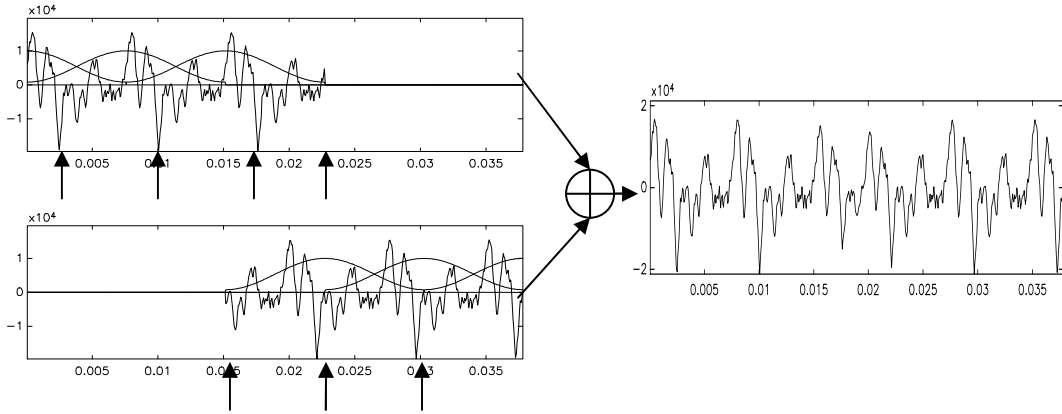
### 5.7.3.1. Parça seçici

“Parça seçici”, ilgili metnin oluşturulabilmesi için veritabanının içerisinde en iyi parça dizisini seçen modüldür. Büyük veritabanlarını kullanıldığı sistemlerde sistem tasarımının en önemli modülüdür çünkü veritabanının etkili bir şekilde kullanılmasını sağlar ve son kalite seçilen parça dizisinin kalitesine (kaliteden kasıt parçaların hedeflenen özelliklere uyması ve ardarda dizildiklerinde sürekli bir işaretin oluşmasının garantilenmesidir) doğrudan bağlıdır. Basit-çalışır sistemimiz veritabanımız ikili ses ve üçlü seslerden oluşmaktadır. Sentezleyicimiz ise ikili ses ekleme yöntemi ile çalıştığından hedef metindeki ikili sesleri veritabanından seçmek zorundadır. Her bir ikili sestem en az bir adet içerdiği için parça seçici sadece hedef ikili seslerle veritabanı ikili sesleri arasında bir seçme yapacak şekilde tasarlanmıştır. Bu işlem yapılırken veritabanındaki ikili ses ve hedef ikili sesin sağındaki ve solundaki ikili sesler de hesaba katılarak en uygun parça seçilmeye çalışılmaktadır. Veritabanında birden fazla ikili ses alternatifi olduğu durumda, hedef ikili ses ve veritabanındaki ikili ses komşulukları arasında bir eşleşme varsa bu durumda eşleşme olan ikili ses seçilmektedir. Birebir eşleşme olmadığı durumlarda ilk sıradaki ikili ses seçilmektedir.

### 5.7.3.2. Sentezleyici modülü: kaynaştırıcı

“Kaynaştırıcı”, seçilmiş işaret parçalarının gerekiyorsa prozodilerinin değiştirilmesi ve kesiklilik oluşmayacak şekilde ardarda dizilmelerini gerçekleştirir. Bu amaçla özellikle 90lı yıllarda birçok işaret işleme yöntemi tasarlanmış ve birbirleriyle çeşitli veriler üzerinden karşılaştırılmıştır. Bu karşılaştırmaların sonucunda gözlenen kalitesi sebebiyle TDPSOLA (Time Domain Pitch Synchronous Overlap Add) [62] isimli algoritma ticari sistemler dahil birçok sistem içerisinde en sık kullanılan algoritmalarından birisi olagelmiştir. Bizim sistemimiz için de TDPSOLA'nın uygun olduğu düşünülmüş ve tasarımın detaylı bir planlaması yapılmıştır. TDPSOLA algoritmasının genel prensipleri kolay olmakla beraber konuşma seslerinin çeşitli karakteristikleriyle uyum içinde çalışabilecek bir şekilde tasarlanması çok sayıda önemli detaylar içermektedir.

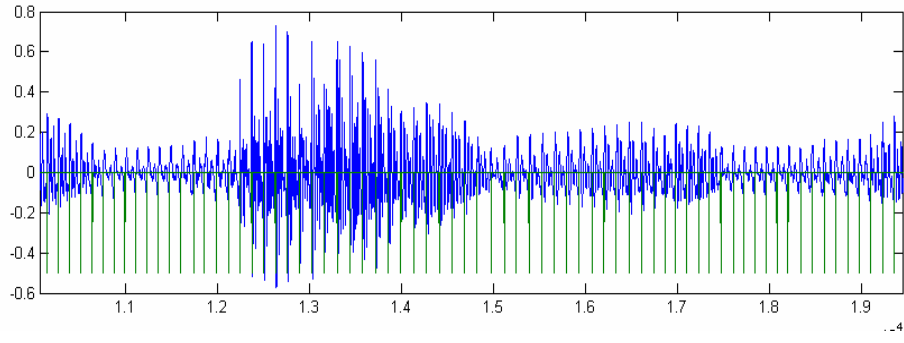
Basit-çalışır sistem için ilk olarak TDPSOLA'nın Şekil 5.10'da gösterilen en basit hali, sadece kaynaştırma işlemi yapan bir modül gerçekleştirilmiştir. Modül verilen iki işareti ardarda kesintisiz eklemek için literatürde “pitch mark” olarak geçen (Şekil 5.11 üzerinde oklarla gösterilmiş) belirli referans noktalarını kullanmaktadır. Bu referans noktaları kullanılarak kaynaştırılacak işaretler küçük bölgelere bölünüp daha sonra bu bölgeler üst üste toplanarak devamlılığı olan bir işaret elde edilmektedir. Bu modülün prozodi değişimini de yapacak şekilde genişletilmesi çalışmaları devam etmektedir.



Şekil 5.10. TDPSOLA kaynaştırıcı

### 5.7.3.3. Veritabanı işaret işleme: TDPSOLA referans noktalarının otomatik tespiti

Veritabanlarının büyüklüğü düşünüldüğünde referans noktalarının tespitinin elle yapılmasının pratik olmadığı açıktır. Bu noktaların tespiti için literatürde çeşitli algoritmalar önerilmektedir. Bu algoritmalar arasında seçim yapmak uzun testler gerektirmekte olduğu için uzman bilgi ve deneyimi ışığında bir seçim yapılmış, ağırlık merkezi değişimine göre referans noktalarını bulan bir algoritma [63] gerçekleştirilmiştir. Gerçeklenen algoritma ses işaretleri üzerinde denenmiş, büyük oranda başarılı olduğu gözlenmiştir.



Şekil 5.11. Konuşma işareti TDPSOLA referans noktaları örneği (aşağı doğru çizgiler referans noktalarının olduğu yerleri göstermektedir)

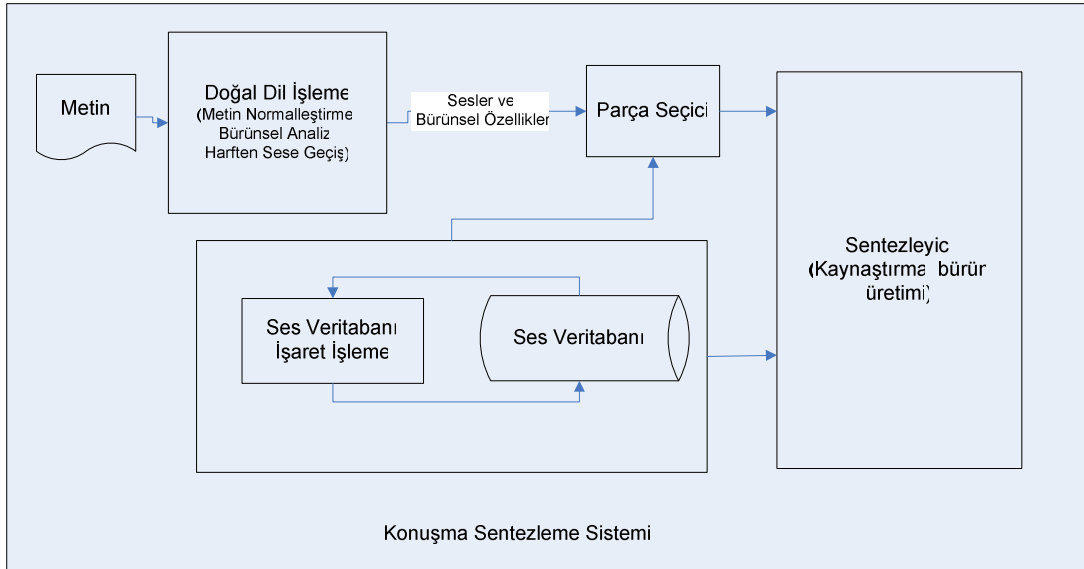
## **BÖLÜM 6. KONUŞMA SENTEZLEME SİSTEM TASARIMI**

Bu bölümde yapılan sistem tasarımı bir MKSS sisteminde bulunması gereken modüller göz önüne alınarak gerçekleştirilmiştir. Bu modüllerin bir kısmı sistemin ana mekanizmalarını, bir kısmı da sentezlenecek konuşma sinyalinin daha kaliteli olmasını sağlayacak destek mekanizmalarını oluşturmaktadır. Bu çalışmada gerçekleştirdiğimiz basit- çalışır sistem modüler bir yapıda olup sadece sistemin ana parçalarından oluşmaktadır. İleride yapılacak çalışmaların sisteme entegre olabilmesi için sistem tasarımı yapılırken tüm modüller hesaba katılmıştır. Geliştirilmesi düşünülen sentezleme sisteminin akış şeması Şekil 6.1'de verilmiştir.

Sisteme entegre edilmeyen modüller harften sese çevirme modülü ve bürün modülleridir. Kullanılan fonem sayısı 29 olarak belirlendiği için harf ve fonemler arasında birebir eşleşme yapılmıştır. Bu yüzden bu modüle ihtiyaç duyulmamıştır. Bu çalışma kapsamında gene sadece akustik modeller üzerine durulmuş ve monoton bir sentez elde edilmeye çalışılmıştır. Türkçe'nin bürünsel özellikleri üzerinde de analizler yapılmış, bu özellikleri temsil eden parametreler (perde frekansı, süre, genlik) sistem tasarlanırken belirlenmiş ama bu parametrelerin aktif olarak nasıl kullanılacağı ile ilgili çok ayrıntılı çalışmalar gerektiğinden, bu çalışma kapsamında kullanılmamıştır.

### **6.1. Kullanıcı Gereksinimleri**

Metinden Konuşma Sentezleme Sistemi Türkçe bir metin dosyasını veya kelime dizisini Türkçe konuşma sinyaline çevirecektir.



Şekil 6.1. Metinden konuşma sentezleme sistemi akış şeması

## 6.2. Sistem Mimarisi

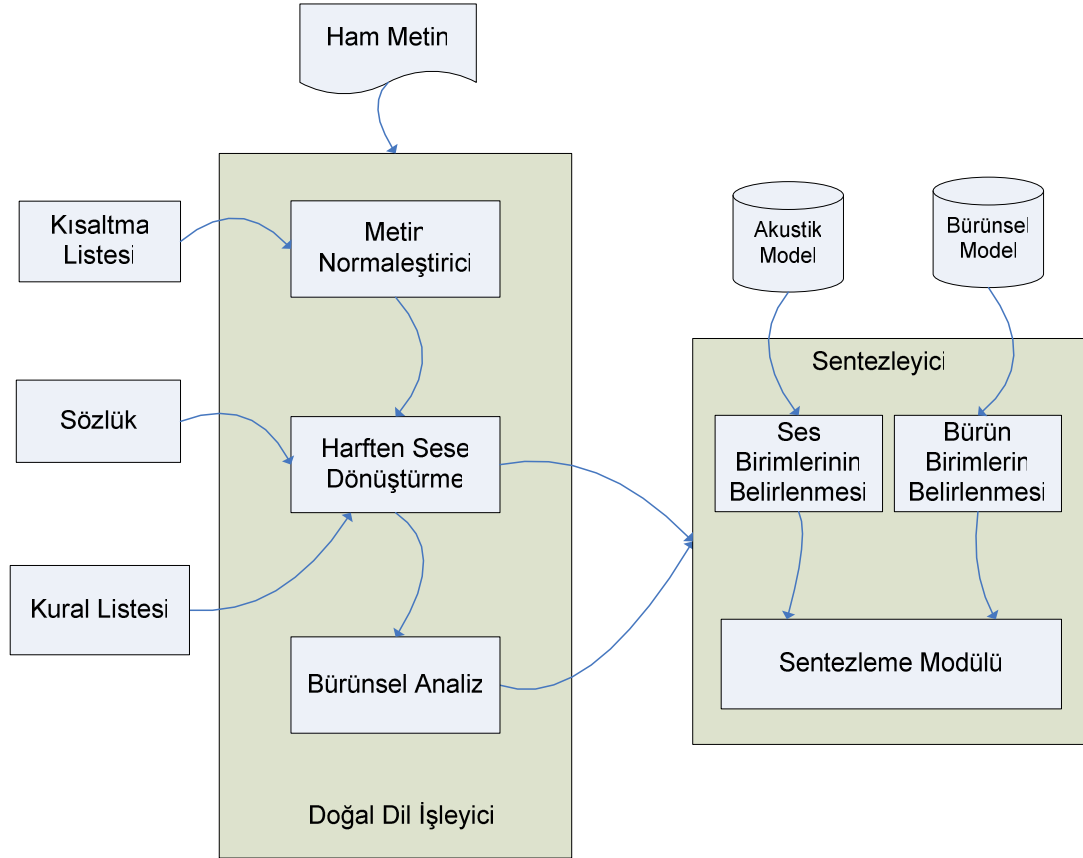
MKS sistemlerinde, verilen metin sentezlenmeden önce, henüz metin aşamasında belirli işlemlerden geçmelidir. Verilen metin önce parçalara ayrılır ve her parça kendi içinde rakam, tarih, kısaltma, sembol gibi standart olmayan kelimeleri içeriyorsa, bu kelimeler tamamıyla harflerden oluşan yazımsal açılımlarına çevrilir. Bu yazımsal açılımlar alfabedeki harflerden oluşmaktadır. Bütün bu aşamalar “Metin Normalleştirme” olarak adlandırılır. Temel birimin ses olduğu sentezleme sistemlerindeki bir sonraki adım, yazımsal ifadeleri bulunan kelimelerin söylemsel ifadelerinin elde edilmesidir. Bu amaçla daha önce elde edilen kurallar veya söyleyiş sözlüğü yardımıyla her kelimenin söyleyişi elde edilir. Bu söyleyiş gene yazılı olarak ifade edilir fakat alfabede olmayan, tanımlanmış özel karakterleri de içerebilir.

“Bürünsel Analiz” modülü süre, perde frekansı, durak, vurgu gibi konuşmanın doğallığını belirleyen etkilerin verilebilmesi için kelimelerin neresinde ve ne kadar ekleneceğinin bilgisini belirler.

Bütün bu bilgiler ışığında sentezleyici “Akustik Model” veritabanından en uygun sesbirimleri seçer ve Bürünsel bilgiler ışığında da “Bürünsel Model” veritabanından



faýdalanarak dođallığı sađlayacak etkileri belirleyerek konuřmayı sentezler. Sistemin genel mimarisi Őekil 6.2 de verilmiřtir.

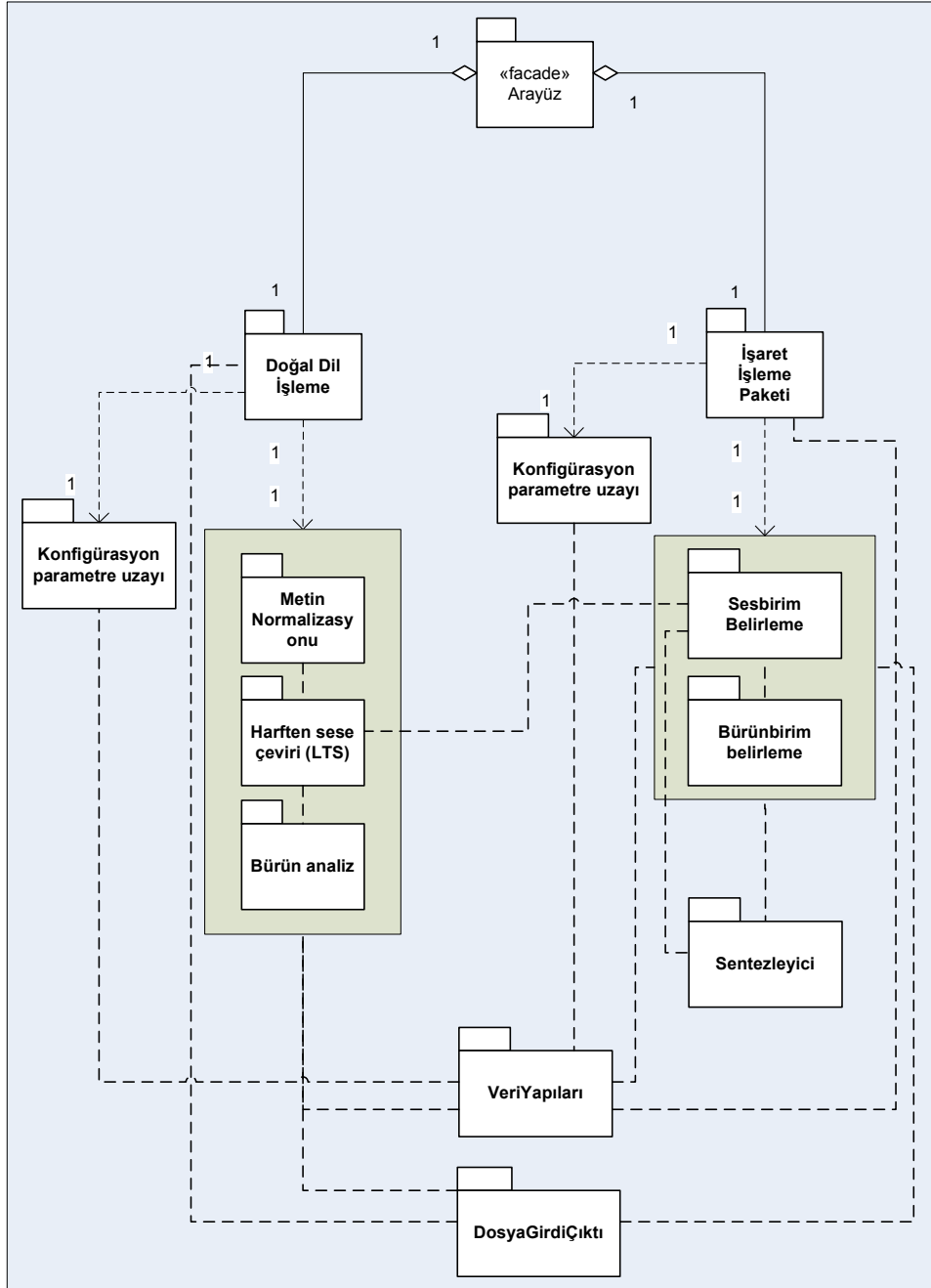


Őekil 6.2. MKS Sistem Mimarisi

### 6.3. Paketler

MKSS, 13 paketten oluřmaktadır. 13 paket arasında 3 ana paket bulunmaktadır: Ara yüz, Tanıma, Dođal Dil İřleme ve İřaret İřleme paketleri. Ara yüz paketi “facade” tasarım řablonunu kullanan ve kullanıcının alt paketlerdeki pek çok ayrıntıda bođulmasını engelleyerek daha rahat bir kullanım sunan bir pakettir. Ara yüz paketi diđer paketlerdeki iřlevleri kullanıcının çok daha basit bir řekilde çağırmasını sađlar. Ayrıca bazı ön kontrolleri yaparak kullanıcının dođru girdiler verdiđinden emin olunmasını sađlar. Dođal dil iřleme paketi, metinden konuřma sentezlemede dilin bürünsel ve fonetik özelliklerinin benzetilmesini ve buna bađlı olarak daha dođal bir sentezleme yapılmasını sađlayan veri yapılarını ve yordamları iřerir. Dođal dil

işleme paketi altında 4 yardımcı paket bulunur. Konfigürasyon parametre uzayı tanımada kullanılan çeşitli parametrelerin yüklendiği bir bellek alanına işaret eder. Bunun haricinde metin normalleştirmeden sorumlu bir paket ile harften-sese çeviri kurallarını ve bürünsel analizleri içeren iki paket daha vardır. İşaret İşleme paketinde de bürünsel ve fonetik özelliklere göre en uygun sesbirimler ve bürün birimler ilgili model veritabanından seçilerek konuşma sentezlenir (işaret düzeyinde). İşaret işleme paketinde sesbirim ve bürün birim belirlemede kullanılan iki paket ile sentezlemeden sorumlu bir paket bulunmaktadır. Bunların yanında Gene işaret işleme paketine bağlı Konfigürasyon parametre uzayı çeşitli parametrelerin yüklendiği bir bellek alanına işaret eder. Bu ana modüller dışında daha genel işleve sahip olan Veri Yapıları Ve Dosya GirdiÇıktı paketleri ise MKSS içinde gerekli olan ve diğer paketler tarafından kullanılan tüm veri yapılarına ait sınıfları içerir ve temel dosya işlemlerini yürütür. MKSS için paket diyagramı Şekil 6.3.'te gösterilmiştir.



Şekil 6.3. MKSS Paket Diyagramı

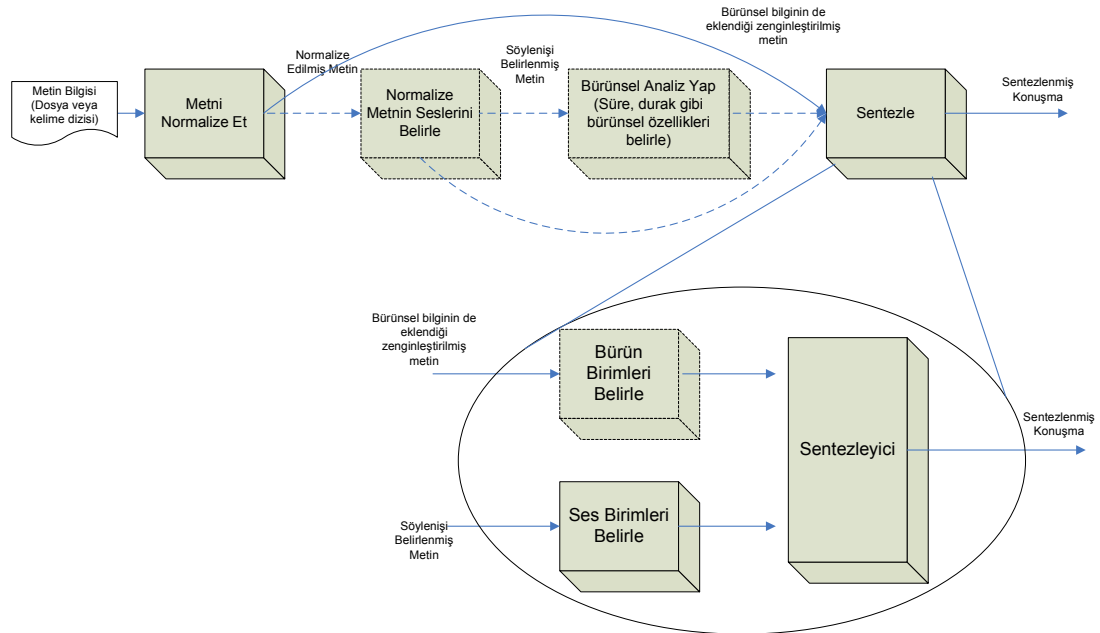
#### 6.4. Veri Akışı

Girdi olarak verilen metin veya kelime dizisi öncelikle “Metin Normalleştirme” modülüne gelir ve çıktı olarak sadece harflerden oluşan ve hiçbir kısaltmanın bulunmadığı normalleştirilmiş bir metin “Harften Sese Dönüştürme” modülüne gönderilir. Sadece harflerden oluşan metin söyleyişine göre tekrar açılır. Bu açılım

esnasında söyleyişteki her ses için farklı bir karakter kullanılır ve söyleyişi belirlenmiş bu metin “Bürünsel Analiz” modülüne gönderilir.Söyleyişi belirlenmiş metin aynı zamanda ilgili sesbirimlerin seçilebilmesi için sentezleyici modülüne de gönderilir.

Bürünsel Analiz modülünde de süre, perde frekansı, durak, vurgu gibi Bürünsel bilgilerin metinde nerede ve nasıl uygulanacağı ile ilgili bilgiler belirlenerek metne eklenir ve elde edilen bu zenginleştirilmiş metin sentezleyiciye gönderilir.

Sentezleyici söyleyişi belirlenmiş metindeki bilgiler yardımıyla akustik model veritabanından uygun ses birimleri seçer ve Bürünsel bilgiler yardımıyla da verilecek Bürünsel etkilerle ilgili modelleri Bürünsel model veritabanından seçerek konuşmayı sentezler. Bunun neticesinde de nihai veri olan konuşma sinyali üretilmiş olur.MKSS sistemi veri akış diyagramı Şekil 6.4 ‘te verilmiştir.



Şekil 6.4. MKSS Veri Akış Diyagramı

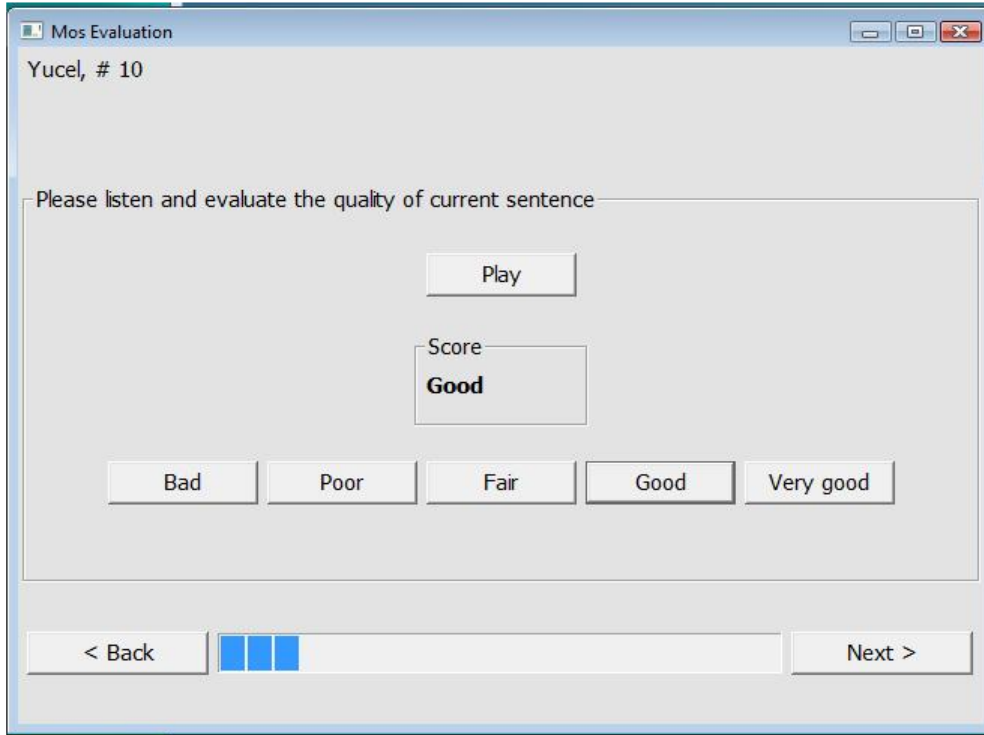
## BÖLÜM 7. DENEYSEL ÇALIŞMALAR

Gerçekleştirilen sentezleyicinin kalitesinin ölçülmesi için MOS (Mean Opinion Score) testleri gerçekleştirilmiştir. Testin öznel olması nedeniyle birden fazla dinleyiciye, çok sayıda sentez çıktısı dinlettirilerek bir skor elde edilmiştir. Testler Şekil 7.1 de arayüz görüntüsü bulunana MOS değerlendirme aracı kullanılmış ve Şekil 7.2'de görülen TÜBİTAK - UEKAE sessiz dinleyici odasında gerçekleştirilmiştir.

100 adet sentezlenmiş cümle, 8 ayrı dinleyiciye 5 farklı oturumda dinlettirilip puanlatılmak suretiyle test gerçekleştirilmiştir. MOS değerlendirme aracı ile, dinleyicilere sentezlenmiş cümleleri dinleterek duydukları sesin kalitesini 5 ölçekli bir skalada (5: En iyi, 1: En kötü) puanlandırmaları sağlanmıştır. Teste tabi tutulanların puanlamaları MOS değerlendirme aracı tarafından saklanmış ve aritmetik ortalamaları alınarak MOS skoru elde edilmiştir. Test için seçilen cümle örnekleri Tablo 7.1'de verilmiştir. Dinlettirilen cümlelerin ortalama uzunluğu 3,7 saniye olarak ölçülmüştür.

Tablo 7.1. Sentezlenen Cümle Örnekleri

Adamcağız bana dua ederek yanıma bindi.
Balkon penceresinde minik bir kedi var.
Ben doğru yolda kaybolmuş kişi görmedim.
Çocuk oynayarak çevresini ve kişileri tanır.
Dayısı Ömer'e iki çift güvercin aldı.
En son hangi kitabı okuduğunu soruyorum.
Fikri hür, vicdanı hür nesiller istiyordu.
Hak bellediğin yolda yalnız gideceksin.
Her akşam dişlerimi yatmadan önce fırçalarım.
Hilmiye teyze çok marifetli ve hamarattır.
Kabul etmeyeceğini nihayet o da anlamıştı.



Şekil 7.1. MOS değerlendirme aracı arayüzü



Şekil 7.2. TÜBİTAK - UEKAE dinleyici odası görüntüsü

Bu test neticesinde elde edilen sonuçlar Tablo 7.2’de gösterilmiştir.

Tablo 7.2. MOS sonuçları

SENTEZLEYİCİ MOS SONUÇLARI										
DİNLEYİCİ	<i>A</i>					<i>B</i>				
TEST ID	<i>A1</i>	<i>A2</i>	<i>A3</i>	<i>A4</i>	<i>A5</i>	<i>B1</i>	<i>B2</i>	<i>B3</i>	<i>B4</i>	<i>B5</i>
TEST MOS	3,82	4	4,22	4,15	4,08	3,97	3,72	3,73	3,71	3,97
DİNLEYİCİ MOS	4,05					3,82				
DİNLEYİCİ	<i>E</i>					<i>F</i>				
TEST ID	<i>E1</i>	<i>E2</i>	<i>E3</i>	<i>E4</i>	<i>E5</i>	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>
TEST MOS	4,23	4,35	4,37	4,08	4,03	4,23	4,28	4,5	4,4	4,14
DİNLEYİCİ MOS	4,21					4,31				
DİNLEYİCİ	<i>C</i>					<i>D</i>				
TEST ID	<i>C1</i>	<i>C2</i>	<i>C3</i>	<i>C4</i>	<i>C5</i>	<i>D1</i>	<i>D2</i>	<i>D3</i>	<i>D4</i>	<i>D5</i>
TEST MOS	3,65	3,59	3,66	3,67	3,62	3,85	3,73	3,73	3,75	3,7
DİNLEYİCİ MOS	3,63					3,75				
DİNLEYİCİ	<i>G</i>					<i>H</i>				
TEST ID	<i>G1</i>	<i>G2</i>	<i>G3</i>	<i>G4</i>	<i>G5</i>	<i>H1</i>	<i>H2</i>	<i>H3</i>	<i>H4</i>	<i>H5</i>
TEST MOS	3,54	3,54	3,53	3,52	3,52	3,81	3,59	3,55	3,59	3,79
DİNLEYİCİ MOS	3,53					3,66				
GENEL MOS	3,87									

## **BÖLÜM 8. SONUÇLAR VE ÖNERİLER**

Bu çalışmada Metinden Konuşma sentezleme sistemleri ile ilgili genel bilgiler verilmeye çalışılmış ve bu sistemlerin Türkçeye uyarlanması için yapılan çalışmalar anlatılmıştır. Hedef dilin Türkçe olması ayrıntılı bir analiz yapılmasını gerektirmiştir. Analiz sonucunda Türkçe için sentezleme sistemlerinde kullanılacak bir ses kümesi belirlenmiştir. Çalışmada basit – çalışır bir sistem geliştirilmesi hedeflendiği için bu ses kümesi yerine Türkçede 29 harf tarafından temsil edilen 29 ses baz alınarak bir ses veritabanı tasarlanmıştır. Bu ses veritabanı logotomlardan oluşacak şekilde ve Türkçe’de en fazla geçen üçlü seslerin istatistiği yapılarak hazırlanmıştır. Daha sonra bu logotomlar kaydedilerek, elle etiketlenmiştir.

Metinden konuşma sentezleme sisteminin gerçekleşmesinde yöntem olarak "Eklemeli Sentezleme" metodu kullanılmıştır. Literatürde sıkça başvurulan "ikili ses sentezleme" ve "birim seçme sentezi" yöntemlerinin temel özelliklerini içeren yeni bir yöntem denenmiştir. İkili ve üçlü seslerin zincirleme eklenmesine dayanan bu yöntemde 3000 adet üçlü ses ve 383 adet ikili ses kaydedilerek bir ses veritabanı oluşturulmuştur. İkili ses veritabanlarına göre daha büyük bir havuzdan, daha büyük parçaların seçilebilmesi, ikili ses sentezinden daha iyi bir sentez elde edilmesini sağlamıştır. Sentezlenmiş konuşmanın kalitesi öznel bir yöntem olan Ortalama Görüş Skoru(Mean Opinion Score-MOS) [64] ölçütü ile değerlendirilmiştir.

Bu çalışma neticesinde Türkçe için ikili-üçlü ses tabanlı, TDPSOLA yöntemini kullanan eklemeli bir sentezleyici geliştirilmiştir. Geliştirilen sentezleyici neticesinde monoton bir konuşma sentezi elde edilmiştir. Bu sayede literatürde verilen ve ticari olarak satılan benzer sistemlerin çıktıları seviyesinde bir sistem gerçekleşmiştir. Bu tip sistemlerin ortak özelliklerinden olan anlaşılabilirliği yüksek, doğallığı düşük bir sistemin daha iyiye taşınması için dilin bürünsel özelliklerini de içeren bir sistemin



hayata geçirilmesinin gerektiği açıktır. Her ne kadar Türkçe'nin bürünsel özellikleri belirlenmiş olsa da bu çalışma kapsamında bürünsel etkiler değerlendirilmemiştir. Bunun temel nedenlerinden bazıları dilbilim literatüründe Türkçenin bürünsel özellikleri konusunda bir uzlaşma bulunmaması, bürünsel modellerin oluşturulmasında yeterli çalışmaların olmaması ve bürünsel etiketleme için gerekli olan verinin hazırlanması sırasında kullanılacak özniteliklerin belirsiz ve öznel olması nedeni ile veri hazırlamada yaşanan zorluklardır. Bürünsel etkilerin sentezleyicilerde kullanılarak daha doğal konuşma üretilmesi bu çalışmanın devamında yapılacak ileri ve yeni çalışmaların temelini oluşturacaktır.

Geliştirilen sistemin daha kaliteli bir konuşma üretmesi için daha büyük ses veritabanlarının tasarlanmasına ve işlenmesi gerekmektedir. Ses veritabanı oluşturulmadan Türkçe'nin ses kümesi dikkate alınarak geniş bir bütüncü belirlenmeli ve bu bütüncü profesyonel konuşmacılar desteğiyle kaydedilmelidir. Ses veritabanı büyüdükçe, bu havuzdan seçilecek hedef parçacıkların seçilmesi süreci de gittikçe karmaşıklaşacaktır. Bu yüzden parça seçici bileşeninde viterbi [65] gibi arama algoritmalarının gerçekleştirilmesi ilerde yapılması gereken çalışmalardır.

## KAYNAKLAR

- [1] DUTOIT, T., High Quality Text- to- Speech Synthesis: An overview, Journal of Electrical & Electronics Engineering, Australia: Special Issue on Speech Recognition and Synthesis, 17, 1, pp. 25-36, 1997.
- [2] DUTOIT, T., V. PAGEL, N. PIERRET, F. BATAILLE, O. VAN DER VRECKEN., The MBROLA Project: Towards a set of high quality speech synthesizers of use for non commercial purposes. The 4th International Conference on Spoken Language Processing, Philadelphia, PA, USA, pp. 1393-1396, 1996.
- [3] KOMINEK, J., BLACK, A. W., CMU ARCTIC databases for speech synthesis. CMU-LTI-03-177. Language Technologies Institute, School of Computer Science, Carnegie Mellon University, 2003.
- [4] ZHANG J., Language Generation and Speech Synthesis in Dialogues for Language Learning, Y.Lisans Tezi, MIT Mühendislik Fakültesi, sf. 54-56, 2004.
- [5] HOLMES, J., HOLMES, W., Speech Synthesis and Recognition, 2nd Edition. CRC, sf. 83-86, New York, 2001.
- [6] LEMMETTY S, Review of Speech Synthesis Technology, Y.Lisans Tezi, Helsinki Üniversitesi, Helsinki, 1999.
- [7] <http://www.festvox.org/history/klatt.html>, Aralık ,2009.
- [8] ALLEN, J., HUNNICUT, S., KLATT, D., Text-to-speech: The MITalk system. Cambridge University Press, sf. 25-31, Cambridge, UK., 1987.
- [9] D. KLATT. Review of text-to-speech conversion for English. Journal of the Acoustical Society of America, 82, 3, pp.737-793, 1987.
- [10] SAGISAKA, Y., KAİKİ, N., IWAHASHİ, N., MİMURA, K., ATR --  
\$nu\$-TALK speech synthesis system. Second International Conference on Spoken Language Processing (ICSLP'92), Banff, Alberta, Canada, volume 1, pp. 483-486, 1992.
- [11] CAMPBELL N., BLACK. A. W., Prosody and the selection of source units for concatenative synthesis. In J. van Santen, R. Sproat, J. Olive, and J. Hirschberg, editors, Progress in speech synthesis, Springer Verlag, pp. 279—282, Berlin, 1996.

- [12] A. HUNT AND A. BLACK. Unit selection in a concatenative speech synthesis system using a large speech database. In ICASSP-96, Atlanta, Georgia, volume 1, pp. 373--376, 1996.
- [13] DUTOIT, T., An introduction to text-to-speech synthesis NonCommercial Purposes. Springer Verlag, pp. 111—132, Berlin, 1996.
- [14] BLACK, A., TAYLOR, P., Building Voices in the Festival Speech System, Language Technologies Institute, School of Computer Science, Carnegie Mellon University, 2000.
- [15] XUEDONG H., FILENO A., HSIAO-WUEN H., MEI-YUH H., KAI-FU LEE AND RONALD R. The SPHINX-II Speech Recognition System: An Overview. Computer, Speech and Language, volume 2, pages 137–148, 1993.
- [16] S. YOUNG, D. KERSHAW, J. ODELL, D. OLLASON, V. VALTCHEV, and P. WOODLAND, The HTK Book. United Kingdom: Entropic Ltd., 1999.
- [17] HARRIS, C., A study of building blocks of speech, The Journal of the Acoustical Society of America, 25, pp. 963-970, 1953.
- [18] KOMINEK, J. AND BENNETT, C.L. AND BLACK, A.W., Evaluating and correcting phoneme segmentation for unit selection synthesis, 8th European Conference on Speech Communication and Technology (EUROSPEECH 2003 - INTERSPEECH 2003), Geneva, Switzerland, pp.313-316, 2003.
- [19] ERGENÇ İ., ÖLMEZ M., Konuşma Dili ve Türkçenin Söyleyiş Sözüğü (Language and Pronunciation Dictionary of Turkish), Türk Dilleri Araştırmaları Dizisi-8, TDK, Ankara, Turkey, 1995.
- [20] T. DUTOIT, et al., The MBROLA project: towards a set of high-quality speech synthesizers free of use for noncommercial purposes, in: Proceedings of the ICSLP'96, 1996.
- [21] DUTOIT, T. AND LEICH, H., MBR-PSOLA: Text-to-speech synthesis based on an MBE re-synthesis of the segments database, Speech Communication, 13, 3-4, pp.435-440, 1993.
- [22] LAMEL L.F., GAUVAIN J.L., PROUTS B., BOUHIER C., BOESCH R., Generation and Synthesis of Broadcast Messages, Proceedings ESCA-NATO Workshop and Applications of Speech Technology, September 1993.
- [23] [http://en.wikipedia.org/wiki/Astro\\_Blaster](http://en.wikipedia.org/wiki/Astro_Blaster), Kasım, 2009.
- [24] [http://en.wikipedia.org/wiki/Space\\_Fury](http://en.wikipedia.org/wiki/Space_Fury), Kasım, 2009.

- [25] [http://en.wikipedia.org/wiki/Star\\_Trek\\_\(arcade\\_game\)](http://en.wikipedia.org/wiki/Star_Trek_(arcade_game)), Kasım, 2009.
- [26] YAMAGISHI, J., ZEN, H., TODA, T. AND TOKUDA, K., "Speaker-Independent HMM-based Speech Synthesis System - HTS-2007 System for the Blizzard Challenge 2007", Proc. Of Blizzard Challenge 2007, 2007.
- [27] DUTOIT, T., GOSELIN, B., "On the use of a hybrid harmonic/stochastic model for TTS synthesis by concatenation", Speech Communication, Elsevier Publisher, 19, pp. 119-143,1996.
- [28] DUBUISSON, T., DUTOIT, T., GOSELIN, B., REMACLE, M., On the Use of the Correlation between Acoustic Descriptors for the Normal/Pathological Voices Discrimination, EURASIP Journal on Advances in Signal Processing, Analysis and Signal Processing of Oesophageal and Pathological Voices, 2009.
- [29] VERONIS, J., HIRST, D., ESSPESSER, R., IDE, N., NL and Speech in Multext project, AAI'94 Workshop on Integration of Natural Language and Speech, 1994.
- [30] DUTOIT, T., CERNAK, M., TTSBOX: A Matlab Toolbox For Teaching Text-To-Speech Synthesis, Proc. ICASSP'05, Philadelphia, USA, 2005.
- [31] BOZKURT B., DUTOIT, T., "An implementation and evaluation of two diphone-based synthesizers for Turkish", Proc. 4th ISCA Tutorial and Research Workshop on Speech Synthesis, pp.247-250, Blair Atholl, Scotland, 2001.
- [32] VURAL, E., A prosodic Turkish Text-To-Speech Synthesizer, Y.Lisans Tezi, Sabancı University Engineering Faculty, 2004.
- [33] SELEN, N., Söyleyiş Sesbilimi, Akustik Sesbilim ve Türkiye Türkçesi, TDK Yayınları, Ankara, 1979.
- [34] TOKLU, O., Dilbilime Giriş, Akçağ Yayınları, Ankara, 2003.
- [35] DEMİRCAN, Ö., Türkiye Türkçesinde Kök Ek Bileşimleri, Ankara, 1977.
- [36] KOCAMAN, A., Dilbilim- Temel Kavramlar, Sorunlar ve Tartışmalar, Dil Derneği Yayınları, Ankara, 2006.
- [37] ERGİN, M., Türk Dil Bilgisi, Bayrak Basım/Yayım/Tanıtım, İstanbul, 2002.
- [38] BANGUOĞLU, T., Türkçenin Grameri, TDK Yayınları, Ankara, 2004.
- [39] ERGENÇ, İ., Konuşma Dili ve Türkçenin Söyleyiş Sözlüğü, Multilingual Yayınları, İstanbul, 2002.

- [40] PALAZ H., BİCİL Y., KANAK A., DOĞAN M.U., “New Turkish Intelligibility Test for Assessing Speech Communication Systems”, Int. Journal of Speech Communication, December 2005.
- [41] ÖZSOY, S., Türkçenin Yapısı –I, Sesbilim, Boğaziçi Üniversitesi yayınları, İstanbul, 2002.
- [42] DOĞAN A., Her Yönüyle dil ve ana çizgileriyle dilbilim, TDK yayınları, Ankara, 1980.
- [43] ÖZKIRIMLI, A., Türk Dili: Dil ve Anlatım (2. Baskı), İstanbul Bilgi Üniversitesi yayınları, İstanbul, 2002.
- [44] GÖKNEL, Y., Modern Türkçe Dilbilgisi, Hür Efe matbaası, İzmir, 1974.
- [45] International Phonetic Association (IPA), Handbook of the International Phonetic Association - A Guide to the Use of the International Phonetic Alphabet, Cambridge University Press, Cambridge, UK, 1999.
- [46] VARDAR, B., Dilbilimin Temel Kavramları ve İlkeleri, TDK Yayınları, Ankara, 1982.
- [47] VARDAR, B., Dilbilim ve Dilbilim Terimleri Sözlüğü, TDK Yayınları, Ankara, 1980.
- [48] SELEN, N., Entonasyon Analizleri, Ankara Üniversitesi Basımevi, Ankara, 1973.
- [49] UNDERHILL, R., “A Lexical Account of Turkish Accent”, Studies on Turkish Linguistics, haz. S. Koç, ODTÜ yayınları, Ankara, 1988.
- [50] BLACK, A.W., SCHULTZ, T., Speaker clustering for multilingual synthesis, Multilingual Speech and Language Processing, Prentice Hall, New York, 2006.
- [51] HUANG X., ACERO A., HON H.W., Spoken Language Processing, Prentice Hall, Upper Saddle River, NJ, 2001.
- [52] SJLANDER, K., AND BESKOW, J. 2000. Wavesurfer - an open source speech tool. In Proc of ICSLP, vol. 4, 464-467.
- [53] SPROAT, R., OLVE, J., An Approach to Text- To- Speech Synthesis, in Speech coding and Synthesis, Amsterdam, pp. 611- 634, Elsevier Science, 1995.
- [54] Türk Dil Kurumu, İmlâ Kılavuzu, Ankara, 1996

- [55] BLACK, A.W., LENZO, K., PAGEL, V., Issues in Building General Letter to Sound Rules, In The Third ESCA Workshop in Speech Synthesis, pp. 77-80, 1998.
- [56] BİCİL Y., DOĞAN M.U., KANAK A., PALAZ H., “Modelling Pronunciation Dictionary of Turkish Language by CART trees,” 14th.IEEE National Conf. Signal Processing and Communication (SIU2006), Antalya, Belek, 2006.
- [57] [http://www.cstr.ed.ac.uk/projects/speech\\_tools/](http://www.cstr.ed.ac.uk/projects/speech_tools/), Kasım, 2009.
- [58] BECKMAN, M. E., HIRSCHBERG, J., SHATTUCK-HUFNAGEL, S., The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (ed.) Prosodic Typology -- The Phonology of Intonation and Phrasing, 2005.
- [59] OSKAY, B., Automatic modeling of Turkish prosody / Türkçe bürün yapısının otomatik modellenmesi, Y. Lisans Tezi, Orta Doğu Teknik Üniversitesi, Fen Bilimleri Enstitüsü, sf. 90 -99, 2002.
- [60] TURK, O., SCHROEDER, M., BOZKURT, B., ARSLAN, L.M., "Voice Quality Interpolation for Emotional Text-to-Speech Synthesis", Proc. Interspeech'05, Lisbon, Portugal, 2005.
- [61] F. SEVERIN, B. BOZKURT, T. DUTOIT, "HNR extraction in voiced speech, oriented towards voice quality analysis", Proc. EUSIPCO'05, Antalya, Turkey, 2005.
- [62] MOULINES E., CHARPENTIER F., “Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones.” Speech Communication 9: 453–467, 1990.
- [63] KAWAHARA H., ATAKE Y., ZOLFAGHARI P., “Accurate vocal event detection method based on a fixed-point analysis of mapping from time to weighted average group delay,” Proc. ICSLP 2000, Beijing, pp. 664–667, 2000.
- [64] ITU-T Recommendation P.800 Methods for subjective determination of transmission quality, 1996.
- [65] FORNEY, G.D., The Viterbi algorithm, Proceedings of the IEEE, 61, 3, pp. 268-278. 1973.

## ÖZGEÇMİŞ

Yücel BİCİL, 16.05.1978 yılında Bitlis'te doğdu. İlk ve orta öğrenimini Bitlis'te, lise öğrenimini Kocaeli'de tamamladı. 1996 yılında başladığı İstanbul Teknik Üniversitesi Kontrol ve Bilgisayar Mühendisliği bölümünü 2001 yılında bitirdi. 2001 yılından itibaren TÜBİTAK-UEKAE'de (Ulusal Elektronik ve Kriptoloji Araştırma Enstitüsü) Çoklu-ortam teknolojileri araştırma ve geliştirme laboratuvarında araştırmacı ve uzman araştırmacı olarak çalıştı. Halen aynı kurumda Uzman araştırmacı olarak görev yapmaktadır. Uzmanlık alanları Türkçe'nin haberleşme testleri, Türkçe'nin fonetiği, metinden konuşma sentezleme sistemleri, gürültü ölçümleri, konuşma sinyalinin iyileştirilmesi ve ses veritabanlarının tasarlanması olarak sıralanabilir. Şu ana kadar rol aldığı bazı projeler şunlardır: NATO dar-bant ses kodlayıcısı, TÜBİTAK - UEKAE Çoklu Ortam Araştırma ve Geliştirme Laboratuvarı Kurulumu, Türkçe'nin Haberleşme Testleri, Türkçe Ses İşleme Platformu projesi, MULTISAUND projesi.