

**T.C.  
FIRAT ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**TÜRKÇE METİNLER İÇİN HECE TABANLI METİNDEN  
KONUŞMA SENTEZLEME SİSTEMİ**

**İlhami SEL**

**Yüksek Lisans Tezi**

**Elektronik ve Bilgisayar Eğitimi Anabilim Dalı**

**Danışman: Yrd. Doç. Dr. Murat KARABATAK**

**TEMMUZ-2013**

**T.C  
FIRAT ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**TÜRKÇE METİNLER İÇİN HECE TABANLI  
METİNDEN KONUŞMA SENTEZLEME SİSTEMİ**

**YÜKSEK LİSANS TEZİ**

**İlhami SEL**

**(102131110)**

**Tezin Enstitüye Verildiği Tarih : 3 Temmuz 2013**

**Tezin Savunulduğu Tarih : 18 Temmuz 2013**

**Tez Danışmanı : Yrd. Doç. Dr. Murat KARABATAK (F.Ü)**

**Diğer Jüri Üyeleri : Doç. Dr. Engin AVCI (F.Ü)**

**Doç. Dr. Yalın Kılıç TÜREL (F.Ü)**

**TEMMUZ-2013**

## ÖNSÖZ

Bu çalışmayı yapmama olanak sağlayan, tez çalışmamın her aşamasında yönlendirmeleri ve güler yüzüyle sürekli destek olan değerli danışman hocam Yrd. Doç. Dr. Murat KARABATAK'a, ayrıca yüksek lisans çalışmalarım boyunca desteklerini benden hiç esirgemeyen değerli hocalarım Doç. Dr. Engin AVCI ve Yrd. Doç. Dr. Davut HANBAY hocalarıma da yardımlarından dolayı teşekkürü bir borç bilirim.

Tüm hayatım boyunca maddi ve manevi desteğini hiç esirgemeyen aileme ve çalışmalarım esnasında sabır, anlayış ve ingilizce çevirileri ile hep yanımda olan hayat arkadaşım Feride SEL'e teşekkürlerimi sunarım.

İlhami SEL  
ELAZIĞ-2013

# İÇİNDEKİLER

## Sayfa No

ÖNSÖZ .....	II
İÇİNDEKİLER.....	III
ÖZET .....	VI
SUMMARY.....	VII
ŞEKİLLER LİSTESİ.....	VIII
TABLolar LİSTESİ.....	X
KISALTMALAR .....	XI
<b>1. GİRİŞ.....</b>	<b>1</b>
1.1 Literatür Araştırması .....	1
1.2 Tezin Amacı.....	3
1.3 Tezin Yapısı.....	3
<b>2. METİNDEN KONUŞMA SENTEZLEME .....</b>	<b>4</b>
2.1 MKS Nedir?.....	4
2.2 MKS Tarihsel Gelişimi.....	5
2.3 MKS Zorlukları.....	7
2.4 MKS Uygulamaları .....	9
2.5 MKS Yöntemleri.....	10
2.5.1 Kural Tabanlı Formant Sentezleyiciler .....	10
2.5.2 Söyleyiş Sentezleyicileri.....	13
2.5.3 Eklemeli Sentezleyiciler .....	14
2.6 Başlıca MKS Sistemleri.....	16
2.6.1 MITALK.....	16
2.6.2 INFOVOX .....	16
2.6.3 BELL LABS TTS SYSTEM .....	16
2.6.4 CNET PSOLA.....	17
2.6.5 ETI ELEQUENCE.....	17
2.6.6 FESTIVAL TTS SYSTEM.....	17
2.6.7 MBROLA .....	17
2.6.8 WHISTLER .....	18

<b>3.</b>	<b>TÜRKÇE METİNDEN KONUŞMA SENTEZLEME SİSTEMİ.....</b>	<b>19</b>
3.1	Türkçenin Dil Ailesi ve Özellikleri.....	19
3.2	Ses, Fon, Fonem ve Hece Kavramları.....	19
3.2.1	Ses, Fonem Ayrımı.....	20
3.2.2	Hece Kavramı .....	22
3.3	Türkçede Harf - Ses İlişkisi .....	23
3.4	Türkçede Bulunan Hece Türleri.....	25
3.5	Yapılan Akademik Çalışmalar .....	26
<b>4.</b>	<b>TÜRKÇE METİNLER İÇİN EKLEMELİ SENTEZLEME YÖNTEMİ.....</b>	<b>28</b>
4.1	Türkçe Metinden Konuşma Sentezleme Yapısı.....	28
4.2	Metin Normalizasyonu .....	28
4.2.1	Kısaltmalar .....	28
4.2.2	Sayılar .....	29
4.2.3	Özel İşaretler .....	30
4.2.4	Dilin Seslendirmedeki Farklılıkları.....	30
4.3	Metnin Hecelere Ayrılması.....	31
4.4	Sesler .....	32
4.4.1	Temel Bilgiler .....	32
4.4.2	Seslerin Ön İşlemden Geçirilmesi.....	33
4.4.2.1	Hanning Pencere Fonksiyonu .....	34
4.5	Seslerin Veritabanına Kaydedilmesi .....	35
4.6	Seslerin Birleştirilmesi .....	37
4.6.1	Örtüşme-Ekleme Algoritması .....	37
4.6.2	Eşzamanlı Örtüşme-Ekleme Algoritması .....	37
4.6.3	Zaman-Alan Senkronize Ton Örtüşme ve Ekleme Algoritması.....	39
4.6.4	Dalga Şekilli- Benzerlik Tabanlı Senkronize Örtüştürme Ve Ekleme Algoritması .....	40
<b>5.</b>	<b>METİNDEN KONUŞMA SENTEZLEME UYGULAMALARI .....</b>	<b>41</b>
5.1	Hece Tabanlı Metinden Konuşma Sentezleme Sistemi .....	42
5.1.1	Seslerin Kaydedilmesi .....	42
5.1.2	Seslerin Data Bölümünün Elde Edilmesi .....	43
5.1.3	Seslerin Ön İşlemden Geçirilmesi.....	46
5.1.4	Seslerin Birleştirilmesi .....	48

5.2	Difon Tabanlı Metinden Konuşma Sentezleme Sistemi .....	51
5.2.1	Çalışma Ortamı .....	51
5.2.2	Uygulama İçeriği.....	52
5.2.3	Metin Seslendirme.....	55
5.2.4	Dosyaların Seslendirilmesi .....	55
5.2.5	Sayı Sistemlerinin Seslendirilmesi.....	56
<b>6.</b>	<b>SONUÇ .....</b>	<b>58</b>
	<b>KAYNAKLAR .....</b>	<b>61</b>
	<b>ÖZGEÇMİŞ.....</b>	<b>65</b>

## ÖZET

Bu tez çalışmasında, Türkçe metinler için iki adet metinden konuşma sentezleme sistemi tasarlanmıştır. Hazırlanan ilk sistemde hece tabanlı olarak kelime sentezi gerçekleştirilmiştir. Oluşturulan ikinci sistem ise nesne tabanlı olarak hazırlanmış ve seslendirilecek parça olarak difon veri tabanını kullanmıştır.

Tasarlanan ilk sistem Matlab alt yapısıyla oluşturulmuş olup eklemeli sentezleme yöntemini kullanmıştır. Bir kelimeyi oluşturan seslerin art arda birleştirilerek seslendirilmesi mantığına dayanan yöntemde birleştirilecek parça olarak heceler seçilmiştir. Bu sistem için hece seslerinin önceden kaydedilmesi ve ötümsüz kısımlarının silinmesi sağlanmıştır. Sonraki işlem olarak sesin Hanning penceresinden geçirilerek diğer seslerle frekanslarının ve enerjilerinin birbirleriyle uyumlu olması sağlanmıştır. Son olarak ses birleştirme algoritmalarından SOLA (Eşzamanlı Örtüşme-Ekleme Algoritması) kullanılarak örnek bir kelimenin sentezi gerçekleştirilmiştir.

Oluşturulan ikinci sistem ise C# programlama dili kullanılarak tasarlanmıştır. Türkçe metin seslendirme konusunda literatürde yer alan birçok özelliğe sahip olan sistem derlenerek Windows işletim sistemlerinde başarıyla kullanılabilir. Sayıların, kısaltmaların ve dosyaların okunması hususunda başarılı olan sistem de noktalama işaretlerinin seslendirmedeki özellikleri de dikkate alınmıştır.

Ayrıca bu tez çalışmasının ilk bölümlerinde ise metinden konuşma sentezleme sistemleri hakkında bilgiler verilmiş, Türkçe dili için hazırlanan sistemler detaylı şekilde incelenmiştir.

**Anahtar Kelimeler:** Metinden Konuşma Sentezleme, MKS, Text to Speech, TTS, Türkçe MKS Sistemleri, Eklemeli Sentezleme Yöntemi, SOLA,

## SUMMARY

### **Syllable Based Text to Speech Synthesis System For Turkish Texts**

In this thesis study two different text to speech synthesis systems are designed. In the first system syllable based word synthesis has been practiced. While the second system has been designed as object based and has utilized diphone database as the articulated part.

Designed first system generated on MATLAB basis has used Concatenative Synthesis Method. In the method which is based on the idea to articulate the phonemes that form the word by joining them in sequence, syllables are selected as the parts to be concatenated. For this system, it is enabled to record syllables beforehand and to delete unvoiced parts. The following process is passing these segments through Hanning Window and adapting these to other segments frequencies and pitches. At last, using one of the phoneme concatenating algorithms, SOLA (Synchronized Overlap Add Algorithm), sample word synthesis is generated.

Second system is generated by using C# programming language. The system that has many specialities consisting in litterateur of articulating Turkish texts can be used successfully in Windows Operating Systems. In the system which is successful in articulating numbers, abbreviations and files, the role of punctuation marks on articulating has also been regarded.

Moreover, in this study's first parts there have been acknowledgements on text to speech systems and systems designed for Turkish language have been examined in detail.

**Key Words:** Text to Speech, TTS, Concatenative Synthesis, Overlap Add, Synchronized Overlap Add Algorithm, SOLA



## ŞEKİLLER LİSTESİ

	<u>Sayfa No</u>
Şekil 2.1.	Kratzenstein'in ses biçimlendirme cihazları..... 5
Şekil 2.2.	Elektronik mekanizmaya sahip ilk konuşma oluşturma sistemi..... 6
Şekil 2.3.	MKS sistemi blok şeması ..... 10
Şekil 2.4.	Formant Frekanslarının LPC analiz filtresinin genlik spektrumundan elde edilmesi ..... 12
Şekil 2.5.	Seri formant sentezleyici ..... 12
Şekil 2.6.	Paralel formant sentezleyici..... 13
Şekil 2.7.	İnsan ses yolunun artikülator modellenmesi..... 13
Şekil 2.8.	Eklemeli sentezleme sistemlerinde bulunan temel adımlar ..... 15
Şekil 3.1.	Hece uzunluk yüzdeleri ..... 26
Şekil 4.1.	Heceleme algoritması ..... 31
Şekil 4.2.	Hanning pencere fonksiyonu akış seması..... 34
Şekil 4.3.	Örnek sinyalin Hanning pencere fonksiyonu uygulandıktan sonraki görüntüsü ..... 35
Şekil 4.4.	OLA metodu ile zaman ölçekleme..... 37
Şekil 4.5.	SOLA metodu ile zaman ölçekleme..... 38
Şekil 4.6.	TD-PSOLA metodu ile zaman ölçekleme ( $\alpha > 1$ ) ..... 39
Şekil 4.7.	WSOLA metodu ile zaman ölçekleme ..... 40
Şekil 5.1.	Hece tabanlı TMKS sisteminin blok şeması..... 42
Şekil 5.2.	“Free Sound Recorder” programı ekran görüntüsü..... 43
Şekil 5.3.	Ötümsüz ve Ötümlü Ses ..... 44
Şekil 5.4.	“il” Sesinin data verisinin elde edilmesi..... 45
Şekil 5.5.	“ha” sesinin data verisinin elde edilmesi..... 45
Şekil 5.6.	“mi” sesinin data verisinin elde edilmesi ..... 46
Şekil 5.7.	Hanning pencereleme yöntemi örnek kodları ..... 46
Şekil 5.8.	“il” Sesinin Hanning penceresinden geçirilmesi..... 47
Şekil 5.9.	“ha” Sesinin Hanning penceresinden geçirilmesi ..... 47
Şekil 5.10.	“mi” Sesinin Hanning penceresinden geçirilmesi..... 48
Şekil 5.11.	SOLA metodu ile zaman ölçekleme..... 49
Şekil 5.12.	Sesin bölüm boyutlarının belirlenmesi için kullanılan kodlar ..... 49
Şekil 5.13.	Sesin bölümlere ayrılması için kullanılan kodlar..... 50

<b>Şekil 5.14.</b>	Sesin bölümlerine efektlerin verilmesi için kullanılan kodlar.....	50
<b>Şekil 5.15.</b>	“il+ha+mi” Hecelerinin PSOLA algoritmasıyla birleştirilmiş halinin grafiksel gösterimi.....	51
<b>Şekil 5.16.</b>	Uygulamanın ekran görüntüsü.....	52
<b>Şekil 5.17.</b>	Konuşmacı seçimi .....	52
<b>Şekil 5.18.</b>	Okuma hızının ayarlanması .....	53
<b>Şekil 5.19.</b>	Dosya menüsü.....	54
<b>Şekil 5.20.</b>	Görünüm menüsü .....	54
<b>Şekil 5.21.</b>	Seslendirilecek dosyanın seçimi .....	55
<b>Şekil 5.22.</b>	Dosyanın seslendirilmesi.....	56
<b>Şekil 5.23.</b>	Sayı okuma modu.....	56
<b>Şekil 5.24.</b>	Telefon numarası seslendirme .....	57
<b>Şekil 5.25.</b>	T.C. kimlik no seslendirme.....	57

## TABLÖLAR LİSTESİ

	<u>Sayfa No</u>
<b>Tablo 3.1.</b> Fonem ayırıcı özellikleri.....	21
<b>Tablo 3.2.</b> Okunuşları ve yazılışları bire bir olmayan istisnai heceler.....	24
<b>Tablo 3.3.</b> Türkçede hecelerin genel yapısı.....	25
<b>Tablo 3.4.</b> Türkçe metinden konuşma sentezleme üzerine yapılan akademik çalışmalar.....	27
<b>Tablo 4.1.</b> Riff veri bölgesi .....	32
<b>Tablo 4.2.</b> Format veri bölgesi .....	33
<b>Tablo 4.3.</b> Data veri bölgesi .....	33

## **KISALTMALAR**

<b>MKS</b>	:Metinden Konuşma Sentezleme
<b>LPC</b>	:Linear Predictive Coding
<b>OLA</b>	:Örtüşme Ekleme Algoritması
<b>SOLA</b>	:Eşzamanlı Örtüşme-Ekleme Algoritması
<b>TTS</b>	:Text to Speech
<b>TMKS</b>	:Türkçe Metinden Konuşma Sentezleme
<b>TD-PSOLA</b>	:Zaman-Alan Senkronize Ton Örtüşme ve Ekleme Algoritması
<b>WSOLA</b>	:Dalga Şekilli- Benzerlik Tabanlı Örtüşürme Ve Ekleme Algoritması
<b>SAPI</b>	: Microsoft System Speech Application

## 1.GİRİŞ

Metinden Konuşma Sentezleme (MKS) yazılı bir metnin elektronik ortamda konuşma sinyallerine dönüştürülme işlemidir. Bu çalışmada Türkçe metinler için hece tabanlı bir metinden konuşma sentezleme sistemi tasarlanmıştır.

Bilgisayarlar ile insanlar arasında konuşma yoluyla iletişim kurma bilgisayar çağının ilk dönemlerinden beri bilim adamları tarafından sıkça çalışılan bir alandır. Bilgisayarların konuşması ve konuşmaları algılaması bilim kurgu filmlerinin de vazgeçilmez bir ögesidir. Bilgisayar bilimleri ve elektronik mühendisliği gibi farklı alanlar bilgisayar konuşmasını gerçekleştirebilmek için çok sayıda çalışma yapmışlardır.

Bu konuda iki temel çalışma alanı vardır. Bunlar; Metinden konuşma sentezleme ve konuşma tanıma sistemleridir. Bunlar farklı şekillerde incelenmektedir. Eğer bu iki çalışma alanı yeterli seviyeye ulaşırsa ileride konuşmayı anlayan ve konuşabilen sistemler geliştirilebilecektir. Hatta bu alanda “Apple” firmasının geliştirmiş olduğu bir uygulama da bulunmaktadır. “Siri” adı verilen bu uygulama sesli komut işlevlerini yapabilmesinin yanında sesli olarak sorulan sorulara da sesli cevap verebilme özelliğine sahiptir.

MKS sistemleri verilen bir metni konuşma sinyallerine dönüştürebilen sistemlerdir. Bu metin kaynağı bir metin belgesi veya elektronik kitap olabileceği gibi bir web sayfası da olabilir. İdeal bir MKS sisteminin amacı insanın okuyabildiği her metni işleyebilmesidir. Tam performanslı bir MKS sistemi sayıları okuyabilmeli, kısaltmaları uygun hale getirebilmeli ve bir kelime için farklı yazımları seslendirebilmelidir.

### 1.1. Literatür Araştırması

Yapılan araştırmalar sonucunda 1779 yılından beri metinden konuşma sentezleme konusunda birçok çalışma olduğu görülmektedir [1]. Zaman içerisinde bu çalışmalar geliştirilerek günümüze kadar devam etmiştir.

Başta İngilizce olmak üzere Türkçe dâhil birçok dil için hazırlanmış MKS sistemleri ticari olarak son kullanıcıya sunulmaktadır [1]. Sunulan ticari sistemler dışında Türkçe Metinden Konuşma Sentezleme (TMKS) alanında birçok akademik çalışmanın da yapıldığı görülmektedir [2].

MKS sistemleri incelendiğinde 3 farklı MKS yöntemi olduğu görülmektedir [3]. Bu yöntemler şu şekildedir:

- Kural Tabanlı Formant Sentezleyiciler
- Söyleyiş (Articulatory) Sentezleyiciler
- Eklemeli (Concatenative) Sentezleyiciler

TMKS alanında yapılan akademik çalışmalar incelendiğinde oluşturulan birçok sistemde eklemeli sentezleme yöntemi kullanıldığı görülmektedir [1], [3-5]. Türkçe dilinin yapısı incelendiğinde eklemeli bir dil olduğu yani kelimelerin hecelerin birleşmesiyle oluştuğu ve bir kelimenin ek olarak farklı kelimelere türeyebildiği görülmektedir [6]. Bu yüzden [1,3,4] çalışmalarında TMKS sistemleri için en uygun yöntemin Eklemeli Sentezleme olduğu belirtilmektedir.

Eklenecek parça olarak bazı çalışmalar [1] ikili üçlü sesleri (difon, trifon) kullanırken bazı çalışmalar [4] ise heceleri kullanmıştır. Bu tez çalışmasında ise iki farklı sistem tasarlanmış hem difon hem de heceler birleştirilecek parça olarak kullanılmıştır.

Eklemeli sentezlemenin kalitesini belirleyen en önemli kısımlar seslerin ön işlemden geçme süreci ve birleştirme yapılırken kullanılan algoritmadır. Birleştirilecek seslerin frekansı birbirleriyle ne kadar uyumlu ise ve birleştirilen kısımların ne kadar uygun bir şekilde örtüştürüldüğü sistemin başarısını doğrudan etkilemektedir. Yapılan araştırmalar sonucunda [7] seslerin birleştirilmesi için 4 farklı algoritma olduğu görülmektedir. Bu algoritmalar şu şekildedir:

- Örtüşme-Ekleme Algoritması (Overlap-Add, OLA)
- SOLA (Eşzamanlı Örtüşme-Ekleme Algoritması)
- TD-PSOLA (Zaman-Alan Senkronize Ton Örtüşme ve Ekleme Algoritması)
- WSOLA (Dalga Şekilli- Benzerlik Tabanlı Senkronize Örtüşürme ve Ekleme Algoritması)

Bu tez çalışmasında literatürde bulunan benzer çalışmalar göz önünde bulundurularak [1,4,8] hece eklemeli sentezleme yöntemi ve SOLA algoritması kullanılmış ve TMKS sistemi oluşturulmuştur.

## 1.2. Tezin Amacı

Bu tez çalışmasının amacı Türkçe dili için tam ve anlaşılabilir bir MKS sistemi oluşturmaktır. Bu amaç doğrultusunda “Matlab” ve “C#” programlama dilleri kullanılarak iki farklı sistem tasarlanmıştır.

Matlab ortamında hazırlanan sistemde eklemeli sentezleme yöntemi kullanılmış ve eklenecek parça olarak heceler seçilmiştir. Wav dosyası şeklinde kaydedilen heceler Hanning Penceresi adı verilen bir dijital filtreden geçirilerek SOLA algoritması kullanılarak birleştirilmiştir.

C# ortamında hazırlanan diğer sistem ise “.Net’in” öğelerinden “Microsoft System Speech Application (SAPI)” kullanılarak oluşturulmuştur. Ses veri tabanı olarak Mbrola projesi ile hazırlanan difon veri tabanı kullanılmıştır. Girilen kelimeleri başarılı bir şekilde seslendiren program kayıtlı metin dosyaları ve farklı sayısal verileri de okuyabilmektedir. Derlenen program farklı ortamlarda da çalıştırılabilmektedir.

## 1.3. Tezin Yapısı

Tezin ikinci bölümünde MKS sistemleri hakkında genel bilgiler verilmiş, tarihsel gelişimi ile MKS sistemleri oluşturulurken karşılaşılan zorluklar hakkında bilgiler verilmiştir. Bunların dışında MKS sistemlerinin nerelerde hangi alanlarda kullanıldığı ve başlıca MKS sistemleri de bu bölümde tanıtılmıştır.

Tezin üçüncü bölümünde Türkçe MKS sistemleri hakkında genel bilgiler ile Türkçe dil ailesinin genel özellikleri verilmiştir. Ayrıca Türkçe dil ailesinin temel öğeleri olan ses, fon, fonem, hece kavramları bu bölümde açıklanmıştır. Türkçede harf-ses ilişkisi, hece türleri ve Türkçe MKS alanında yapılan akademik çalışmalar da aynı şekilde bu bölümde açıklanmıştır.

Dördüncü bölümde Türkçe metinler için hece tabanlı bir metinden konuşma sentezleme sisteminin nasıl oluşturulabileceği açıklanmıştır. Metin normalizasyonu sırasında hangi işlemlerin yapılabileceği ve metnin hecelere nasıl ayrılabilceği bu bölümde belirtilmiştir. Temel ses bilgileri, hecelerin kaydedilmesi, seslerin önışlemden geçirilmesi ve seslerin birleştirilmesini sağlayan algoritmalar da bu bölümde açıklanmıştır. Beşinci ve son bölümde ise hazırlanan uygulamalar ve işlem basamakları belirtilmiştir.

## 2. METİNDEN KONUŞMA SENTEZLEME

### 2.1. MKS Nedir?

Metinden Konuşma Sentezleme; bilgisayar tarafından yazılı bir metnin ses sinyallerine dönüştürülme işlemidir. Günümüzde MKS uygulamaları çoklu ortam araçlarında kullanıcı ile olan etkileşimin artırılması için kullanılmaktadır. Görme engelliler başta olmak üzere belirli kitleler için geliştirilen bu sistemler sabit veya taşınabilir birçok cihaz üzerinde kullanılabilen bir teknolojidir.

Bir MKS sistemi başlıca iki bölümden oluşur. Bunlar metin işleme ve konuşma sentezleme bölümleridir. Metin işleme bölümü girdi metni ayrıştırarak konuşma sentezleme bölümüne hazırlamakla sorumludur. İdeal bir sistemde metin işleme bölümü oldukça karmaşık olabilir. Çünkü her metni doğru bir şekilde işleyebilmek ve doğru sonuç üretebilmek oldukça zor bir süreçtir.

Metin işleme bölümünün esas amacı girdi metni konuşma sentezleme bölümü tarafından işlenebilecek doğru alt bölümlere ayırabilmektir. Doğru alt parçalar konuşma sentezleme bölümünde kullanılan sentezleme tekniğine göre değişebilir. Örneğin hece tabanlı olarak hazırlanmış eklemeli yöntemlerde metin işleme bölümü girilen metni hecelere, difon tabanlı hazırlanan sistemlerde ise metinleri difonlara ayırmalıdır. Metin işleme bölümünün diğer görevleri ise doğru tonlamaya karar verebilme, kısaltmaları çevirebilme ve sayıları okunabilecek birimlere çevirebilmektir.

Konuşma sentezleme bölümü çıkış verisini yani son konuşmayı sentezlemeden sorumludur. Kural tabanlı ve birleştirilmiş sentez olmak üzere konuşma sentezleme tekniğinin iki popüler çeşidi vardır [1]. Kullanılan sentezleme tekniğinin çeşidine göre sistem için bazı önlemlerde yapılmalıdır. Örneğin birleştirilmiş sentezleme yöntemi kullanılacaksa ses veri tabanının hazırlanması ve kaydedilen seslerin düzenlenmesi gibi işlemler önceden yapılmalıdır.

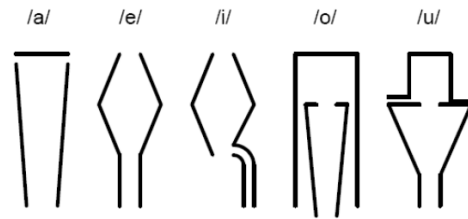


Bu basamaklar sonuçlandırıldığında doğru ve kararlı bir şekilde çalıştırılabilen bir MKS sistemi oluşturulabilir. Bu şekilde oluşturulabilecek bir MKS sistemi oldukça geniş bir kullanım alanına sahip olacaktır.

Hali hazırda oluşturulmuş MKS sistemlerinin günümüzde çok yaygın kullanım alanları bulunmaktadır. Bunlardan bir kaçısı sesli yanıt sistemleri, bilgi ve uyarı sistemleri, görme engelli kişiler için sözlü iletişim aracı vb. gibidir. Sonuç olarak bu şekilde oluşturulan bir sistemle dijital ortamda bulunan bütün yazıların sesli olarak okutulması mümkün olmaktadır.

## 2.2. MKS Tarihsel Gelişimi

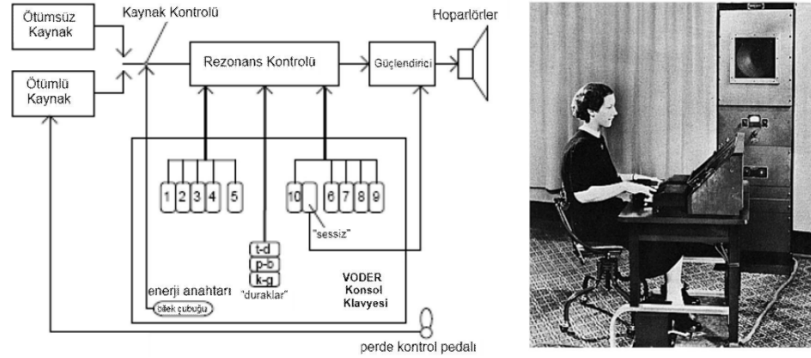
Elektronik sinyal işleme icat edilmeden çok önceleri insan sesi üretebilen makineler yapılmaya çalışılmıştır. Bu makinelerin ilk örneği Aurillac'lı Gerbert, Albertus Magnus ve Roger Bacon'ın içinde yer aldığı "Konuşan Kafalar" isimli çalışmadır [1]. 1779 yılında Rus Bilimler Akademisi'nde çalışan Danimarkalı bilim adamı Christian Kratzenstein beş uzun sesli harf sesi üretebilen insan ses yolu modelleri inşa etmiştir (Şekil 2.1). Uluslararası Dil Bilimi Alfabeti Gösterim Sistemi'ne göre bu harfler (a, e, i, o, u)'dur [9,10].



Şekil 2.1. Kratzenstein'in ses biçimlendirme cihazları [11].

Bunun ardından Macaristan Pressburg'dan Wolfgang von Kempelen 1791 tarihli bir çalışmasında körükle çalıştırılan "Akustik-Mekanik Konuşma Makinesi" geliştirmiştir [10,12]. Bu makinede dil ve dudakların da modelleri eklenerek sessiz harflerin de seslendirilmesi sağlanmıştır. 1837 yılında Charles Wheatstone von Kempelen'in tasarımı üzerine bir "Konuşan Makine" icat etmiş ve 1857'de M. Faber "Euphonia"yı yapmıştır. Wheatstone'un tasarımı 1923 yılında Paget tarafından tekrar canlandırılmıştır [1,9].

1930’larda Bell Laboratuvarlarında çalışan bilim adamları konuşma sesinin temel ton ve rezonanslarını otomatik olarak analiz eden “Vocoder” cihazını geliştirmişlerdir. Vocoder ile ilgili çalışmalarının üzerine Homer Dudley, 1939 New York Dünya Fuarı’nda sergilediği ve elle çalıştırılan “Voder” isimli ses sentezleyicisini geliştirmiştir (Şekil 2.2) [1,12,13].



Şekil 2.2. Elektronik mekanizmaya sahip ilk konuşma oluşturma sistemi [13].

“Desen Geri Çalımı” Dr. Franklin S. Cooper ve meslektaşları tarafından 1940’ların sonuna doğru imal edilmiş ve imalat 1950’de tamamlanmıştır. Bu donanımın birkaç sürümü olsa da günümüze sadece bir tane kalmıştır. Makine sesin akustik desenlerine ait resimleri bir spektogram şeklinde yeniden sese dönüştürmektedir. Alvin Liberman ve meslektaşları da bu cihazı kullanarak ses bilimsel parçacıklara (sessiz ve sesli harfler) dair akustik ipuçları keşfetmişlerdir [1].

1980 ve 1990’lı yıllarda etkin olan sistemler daha çok MIT’den Dennis Klatt’ın çalışmaları üzerine temellendirilmiş MITalk ve Bell Laboratuvarları’nın sistemleridir. Bell’in sistemi doğal dil işleme yöntemlerini sıkça kullanan ilk çok dilli lisandan-bağımsız sistemlerden biridir [1].

İlk bilgisayar temelli ses sentezleme sistemleri 1950’lerin sonunda üretilmiştir. İngilterede kullanılan ilk metinden sese dönüştürme sistemi Noriko Umeda ve meslektaşları tarafından 1968 yılında Japonya’daki Elektroteknik Laboratuvarı’nda geliştirilmiştir [1,10,14]. 1961 yılında fizikçi John Larry Kelly, Jr. Ve meslektaşı Louis Gerstman bir IBM 704 bilgisayar kullanarak Bell Laboratuvarları tarihinin en önemli olaylarından biri haline gelen bir işe imza atarak ses sentezlemişlerdir. Kelly’nin ses kaydedici sentezleyicisi (vocoder) Max Mathews’un müzikal eşliği ile “Daisy Bell” şarkısını seslendirmiştir. Rastlantı eseri Arthur C. Clarke arkadaşı ve meslektaşı olan Bell Laboratuvarları Murray Hill tesisinde çalışan John Pierce’ı ziyaret eder ve bu ürüne ait

sunumdan o kadar etkilenir ki ürünü 2001 yılında çektiği “A Space Odyssey” adlı romanın film versiyonundaki doruk noktasını oluşturan sahnede kullanır [1].

Tamamen elektronik ses sentezlemenin başarısına rağmen mekanik ses sentezleyicilerine yönelik araştırmalar hala devam etmektedir. 1970’lerde ses sentezleme özellikli Talker Serisi Antropomorfik Konuşan Robot Waseda ortaya çıkmış, ilk örneklerden birisi de 1976’da Telesensory Systems Incin (TSI) görme engelliler için çıkardığı “Speech+” taşınabilir hesap makinesi olmuştur [1]. Benzer dönemlerde Texas Instruments’ın ürettiği “Speak&Spell” gibi öncelikli olarak eğitim amaçlı üretilen cihazlar yapılmıştır [1]. 1979 yılında “Fidelity” elektronik satranç bilgisayarının konuşan bir sürümünü çıkartırken 1980 yılında ses sentezlemeyi ilk kullanan video oyunu “Stratovox” Sun Electronics tarafından sunulmuştur [1,14].

Tüm bunlar ele alındığında ilk elektronik ses sentezleyicileri robot sesi olarak bildiğimiz türden sesler üretip, çoğu zaman zar zor anlaşılabilir bir sonuç vermekteydi. O günden bu yana sentezlenmiş ses kalitesi devamlı olarak gelişse de çağdaş ses sentezleme sistemlerinden alınan ses çıktıları hala gerçek insan sesinden rahatlıkla ayırt edilebilmektedir.

### **2.3. MKS Zorlukları**

Bir bilgisayar programından beklenen karşılaşılabileceği olası her durum için doğru bir şekilde geri dönüt verebilmesidir. Bunun için bir bilgisayar o durum doğrultusunda programlanmalıdır. Programcı programın karşılaşılabileceği durumları öngörebilmelidir. Bu bazı durumlar için kolay iken, doğal konuşma işleme gibi bazı durumlar da çok zor olabilir. Sisteme girilebilecek her olası metni belirlemek çok zor olduğu için klasik programlama yaklaşımlarından farklı bazı teknikler kullanılmalıdır.

Bir MKS sistemi doğal bir dil metnini ele alır bu yüzden bir MKS sisteminin çok değişik problemlerle karşılaşması beklenmektedir. Bunlar kısaltmalar, sayılar, noktalama işaretleri vurgu ve tonlamalar olabileceği gibi telaffuzda olabilmektedir. Eğer bir dil fonetik değilse yani yazıldığı gibi okunmuyorsa MKS sistemine telaffuzlarını da kaydetmek bir çözümdür. Ancak bu hem kayıt işlemi hem de hafıza açısından maliyetli bir çözümdür.

Diğer bir çözüm telaffuz için genel kurallar üretmek ve bu kuralları girdi kelimelere uygulamaktır. Bu çözüm programlama mantığı açısından daha iyi bir çözümdür ancak çok

iyi bir dilbilimsel araştırma gerektirir. Çünkü her kural tüm kelimelere uygulanamayabileceği için bazı istisnai durumlarda da başarısız olabilir [1].

Telaffuzla ilgili başka bir problem ise telaffuzdaki belirsizliktir. Bir kelime için farklı telaffuzlar mümkün olabilir ve metin işleme bölümü hangisinin doğru olduğuna karar verebilmelidir [2].

Konuşma sentezlemede ki başka bir zorluk ise tonlamadır. Tonlamanın yerine göre içerik ve anlam değişebileceğinden dolayı doğru tonlama çok önemlidir. Aşağıda verilen iki cümle bu duruma örnek gösterilebilir.

- -Kim su içmek ister?                      -Ben su içmek isterim
- -Ne içmek istersin?                      -Ben su içmek isterim.

Tonun kullanımı kişiye, onun ruhsal durumuna ve genel yapısına bağlı olarak çeşitlilik gösterebildiği gibi ulusların karakterlerine uygun olarak değişik biçimlerde ortaya çıkabilir [16]. Örneğin İtalyanca ve İspanyolca, kullanılan ton yüksekliği açısından önde gelirken, Türkçede zayıf, İngilizcede ise çok zayıf ton yüksekliklerinden söz edilebilir [17].

Metni işleme bölümü konuşma sentezleme problemlerinin sadece bir kısmıdır. Doğru telaffuz ve tonlamaya karar verdikten sonra konuşma sentezleme bölümü bunu tanıyabilmelidir. Bu çok zor bir görevdir. Çünkü algısal olarak aynı sesler farklı içerikte akustik olarak farklı olabilir [1]. Örneğin, “bana” kelimesiyle “banane” kelimesindeki heceler aynı olmasına rağmen okunuşu farklı olmaktadır.

Bir sesin kelimedeki yeri veya vurgulanıp vurgulanmadığı gibi durumlar o sesin tam süre ve frekanslarını doğrudan etkilemektedir. Metin işleme bölümü tonlamayla da ilgilendiği için tonlamanın doğal bir konuşma için nerede olacağına karar vermelidir. Ancak bunu algılamak konuşma sentezi bölümünün görevidir. Tonlama mekanizması henüz tam olarak çözülememektedir. Farklı tonlama modelleri bulunmaktadır ancak bunların hiçbiri her durumda doğru bir şekilde çalışma konusunda yeterince başarılı değildir [1].

Tonlama dışında sayıları okumakta MKS sistemi için zor bir görevdir. Sistem önce ne tür bir numara olduğunu anlamalı ve buna uygun şekilde davranmalıdır. Telefon numaraları, T.C. kimlik numaraları, normal numaralardan farklı okunur. Bunun dışında eğer bir markanın seri numarasıysa da okunuşu farklı olacaktır. Bunların her biri için ayrı okuma formatı geliştirilmesi gerekmektedir.

Metin işleme bölümü kısaltmalarla da baş edebilmelidir. Günlük hayata dile her geçen gün yeni kısaltmalar eklendiği için tüm kısaltmaları içerisinde barındırabilecek bir sistem oluşturmak çok zordur.

İdeal bir MKS sistemi tüm bu problemler için çözümler üretmelidir. Henüz bu problemleri mükemmel bir şekilde çözebilecek bir sistem bulunmamaktadır. Bütün sistemler en iyisini yapmaya çalışmaktadırlar. Bu problemler tam anlamıyla çözüme ulaştırıldığında oluşturulabilecek bir MKS sistemi ideal veya ideale yakın bir sistem olacaktır. Yani bu problemleri çözümlemede ki başarı seviyesi sistemin kalitesini doğrudan etkilemektedir.

## **2.4. MKS Uygulamaları**

Bilgisayarlar şuan ki kadar yaygınlaşmadan önce bir şeyler öğrenmek veya araştırmak için dergi, kitap veya benzeri şeyleri okumak en sık başvurulan yöntemlerin başında geliyordu. Sonradan internetin yaygınlaşmasıyla beraber bilgisayarda en büyük araştırma ve bilgiye erişme kaynağı haline gelmiştir. Bu açıdan bakınca MKS sistemi dolaylı olarak bilgiye erişmede birçok kişinin ihtiyacını giderebilecek bir sistemdir.

Şüphesiz MKS sistemleri görme engelli insanlara bilgisayar dünyasına girişlerinde en büyük kolaylığı sağlayan sistemlerdir. Bilgisayarı kullanmalarını kolaylaştırdığı gibi her metni okuyabilir ve yazdıklarını kontrol edebilirler. İnternetin metin bölümlerine erişip normal bir kullanıcının kullandığı birçok uygulamayı kullanabilirler.

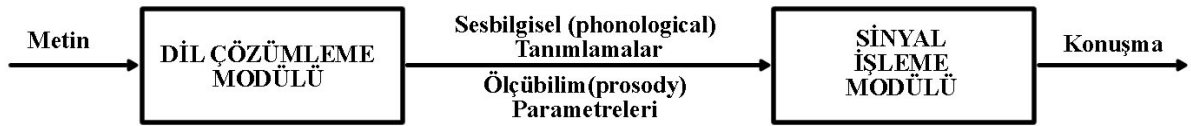
MKS'nin bir diğer kullanımı telefon bağlantısı aracılığıyla bilgisayara uzaktan erişimdir. Bu rezervasyon şirketleri tarafından sıkça kullanılan bir yöntemdir. Bankacılık ve finans şirketleri bu teknolojiyle hesap bilgilerini kullanıcıya sağlamada veya telefon hattıyla yeni işlemler yapmada kullanabilirler [1]. Böylece insanlar bankaya gitmek ve basit işlemler için sıra beklemek zorunda kalmazlar. Bu teknolojinin en güzel yanı özel bir donanım gerektirmemesidir. Dünyanın herhangi bir yerinde bulunan telefon bankaya kolaylıkla ulaşabilir.

Sentezlenen konuşma birçok eğitimsel durumda da kullanılabilir. Konuşma sentezleyicisi bulunan bir bilgisayar birçok farklı dilin yazılışını ve telaffuzunu öğretmek için programlanabilir.

Günümüz teknolojisinde e-posta iletişimi önemli bir yer almaktadır. Ancak bazı durumlarda maillere ulaşmak çok zor hale gelebilmektedir. Bu tür durumları aşmak için mailleri okuyabilen bazı sistemler de hali hazırda bulunmaktadır.

## 2.5. MKS Yöntemleri

Metinden konuşma sentezleme sistemleri genel olarak iki ana bölümden oluşmaktadır. Bu şekilde hazırlanan sistemlerin temel blok şeması Şekil 2.3'te verildiği gibidir [3]. Bu bölümlerden biri dil çözümüleme modülü (metin işleme) diğeri ise sinyal işleme modülü (konuşma sentezleme) olarak adlandırılırlar [3]. Dil çözümüleme modülünde konuşmaya çevrilecek metin bazı ön işlemlerden geçirilir. Bu sayede metinden sentezlenecek konuşmayı oluşturan ses bilgilerinin elde edilmesi sağlanmaktadır. Eğer metinde kısaltmalar varsa ('Dr.', 'T.C.', 'Mah.' gibi) bunlar okunuşlarındaki hallerine çevrilmelidir [3]. Ayrıca sayılar ("tarihler", "1. gibi kısaltmalar" vb.) da okunabilir bir metne dönüştürülmelidir. Sinyal işleme modülü ise; sesbilgisel ve bürünsel (ölçü bilim) bilgileri kullanarak konuşmayı oluşturur [3].



Şekil 2.3. MKS sistemi blok şeması [3].

Metinden Konuşma Sentezleme Sistemlerinde 3 yaklaşım görülmektedir. Bu yaklaşımlar şunlardır;

1. Kural Tabanlı Formant Sentezleyiciler
2. Söyleyiş (Articulatory) Sentezleyiciler
3. Eklemeli (Concatenative) Sentezleyiciler

### 2.5.1. Kural Tabanlı Formant Sentezleyiciler

Bu tip Sentezleyiciler konuşma sinyalinin doğrusal ön görümlü kodlaması (LPC, Linear Predictive Coding) temeline dayanmaktadır. Yani konuşma sinyalinin n. örneğinin  $s(n)$ ,

önceki p adet örneğinin doğrusal kombinasyonu şeklinde ifade edilebilir. Bu durum (Eşitlik 2.1) de gösterilmektedir [3].

$$\hat{s}(n) = -\sum_{i=1}^p a_i s(n-i) \quad (2.1)$$

Konuşmayı; durağan kabul edildiği 20-30 ms'lik kısa sürelerde oldukça başarılı bir şekilde kodlayabilen bu yöntemde  $a_i$ 'ler; LPC katsayıları olarak adlandırılır ve bu süre zarfında (çerçeve süresi) sabit olmaktadır [3]. Konuşmanın gerçek değeri ile öngörülen değer arasındaki fark; hata sinyalidir ve sentezlemede uyarım sinyali olarak adlandırılmaktadır (Eşitlik 2.2) [3].

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{i=1}^p a_i s(n-i) \quad (2.2)$$

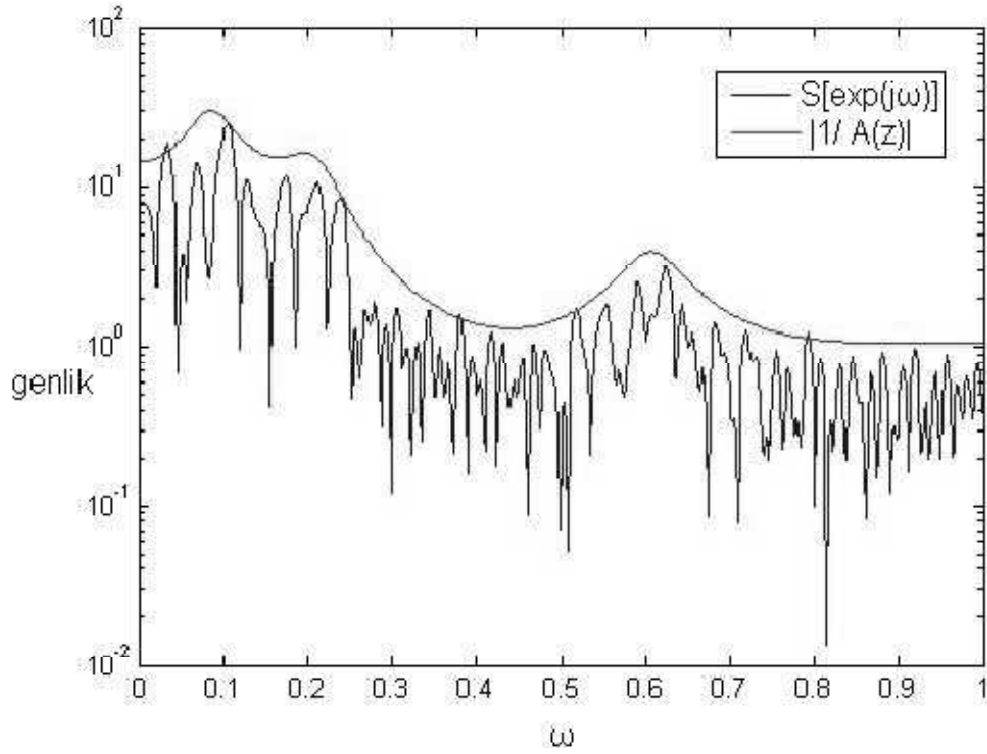
Hata sinyali ile birlikte LPC eşitliği (Eşitlik 2.3'de) gösterildiği gibi olacaktır.

$$s(n) = -\sum_{i=1}^p a_i s(n-i) + e(n) \quad (2.3)$$

Eşitlik 2.3'ten z-dönüşümü ile aktarım işlemi (transfer fonksiyonu) hesaplanırsa sonlu uzunlukta dürtü tepkili bir süzgeç elde edilmektedir:  $A(z)$ -LPC analiz filtresi (Eşitlik 2.4) süzgeç katsayıları  $a_i$ ; hatanın karesini minimum yapacak şekilde hesaplanır ve  $A_z$  süzgeci; ses yolunu (vocaltract) modeller. Böylece konuşma sinyali, incelendiği zaman aralığında, uyarım sinyaline ve süzgeç katsayılarına ( $a_i$ ) ayrıştırılmış olur [3].

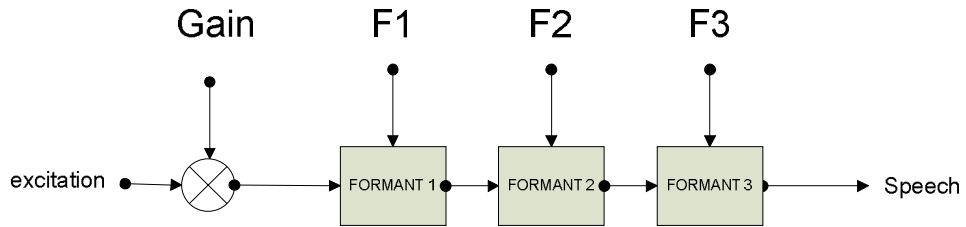
$$\frac{E(z)}{S(z)} = A(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (2.4)$$

Analiz süzgecinin genlik tepkisinde yer alan tepeler, ses yolunun rezonans frekanslarına karşılık gelir ve formant frekansları olarak adlandırılmaktadır (Şekil 2.4). Bu sebeple formant sentezleyiciler; konuşmayı bu ana bileşenleri (formant frekanslarını ve bant genişliklerini) ayarlayarak sentezlemeye çalışmaktadırlar. Tarihsel olarak en eski geliştirilen sentezleyiciler formant sentezleyicilerdir [3].



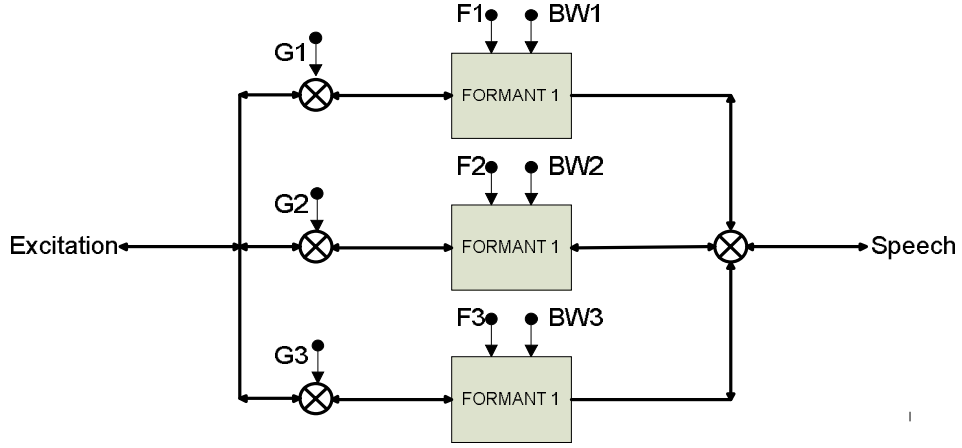
Şekil 2.4. Formant Frekanslarının LPC analiz filtresinin genlik spektrumundan elde edilmesi [3].

Seri formant sentezleyiciler, denetim bilgisi olarak sadece formant frekanslarına ihtiyaç duymaktadırlar (Şekil 2.5). Ünlülerin sentezinde, formant frekanslarındaki genlik değerlerini ayrı ayrı kontrol etmeye gerek yoktur [3]. Seri formant sentezleyicilerin sürtünücü ve patlamalı ünsüzleri sentezlemede çok başarılı olmadığı belirtilmektedir [17]. Paralel yapıya göre daha az denetim bilgisine sahip olduğu için daha kolay gerçekleştirilebilir.



Şekil 2.5. Seri formant sentezleyici



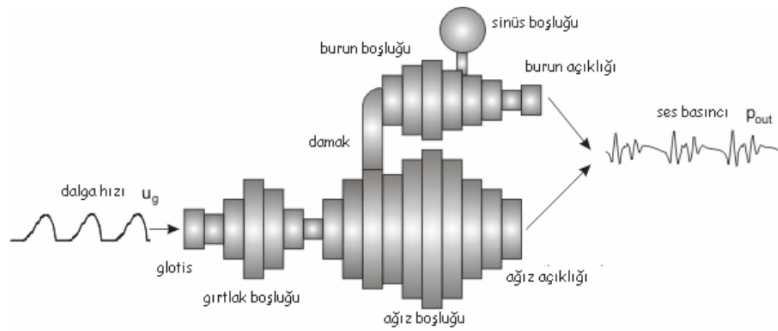


Şekil 2.6. Paralel formantsentezleyici

Paralel formant sentezleyiciler, her bir formant frekansı için ayrı frekans, genlik ve bant genişliği parametrelerine ihtiyaç duymaktadır (Şekil 2.6). Dolayısıyla daha karmaşık yapıdadır. Geniz (nasal) ünsüzlerinin ve sürtünücü ünsüzlerin sentezinde daha iyi sonuçlar verdiği, ancak ünlülerin sentezini gerçekleştiremediği belirtilmektedir [17].

### 2.5.2. Söyleyiş Sentezleyicileri

Söyleyiş (Articulatory) sentezleyicilerinde amaç; tüm ses birimlerin (phoneme), insanın ses üretim mekanizmasında nasıl oluşturulduğunun en hassas şekilde modellenmesidir [3]. Şekil 2.7’de örnek bir söyleyiş sentezleyici gösterilmektedir [15, 18].



Şekil 2.7. İnsan ses yolunun artikülör modellenmesi [15,18].

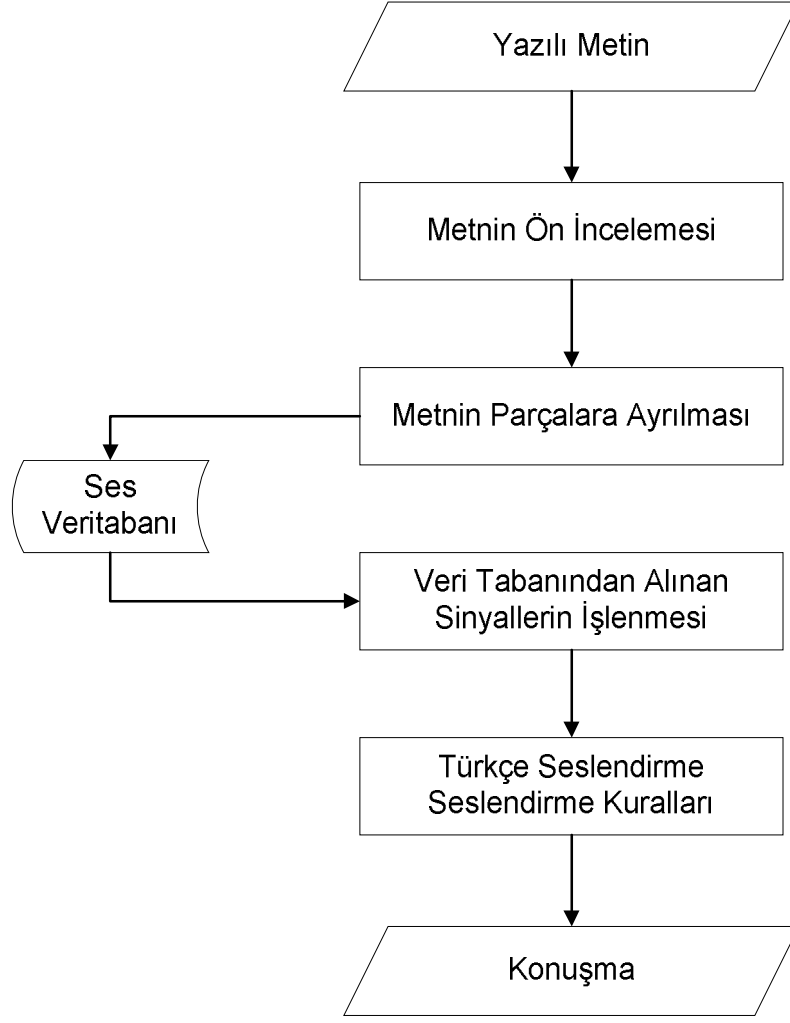
Teorik olarak en doğru yaklaşım söyleyiş sentezleyicilerdir; fakat ses yolunda bulunan organların davranışlarını modellemek pek de kolay değildir [3]. Çünkü bu organların (ses telleri, ağız ve geniz boşlukları, dil, dudaklar vd.) görüntülerini elde etmek zor olmaktadır.

Tahmin edileceđi üzere çok sayıda parametre ieren bu modelin elde edebildiđi konuřma kalitesi zellikle konuřmanın geiř blmlerinde, formant sentezleyicilerinkinden daha iyidir fakat iřlem dnleřimi (tradeoff) de bir o kadar fazladır [17].

### **2.5.3. Eklemeli Sentezleyiciler**

Eklemeli (Concatenative) Sentezleyiciler dođal sesleri kaydedip, bunları birleřtirerek ses sentezlemek, anlaşılır sesler retmenin en kolay yoludur. Fakat bu kolaylıđa karřın, genellikle bir konuřmacı ve bir sesle sınırlanması, diđer yntemlere gre daha fazla bellek kapasitesi istemesi gibi olumsuzluklar, seslerin birleřtirerek ses retmenin dezavantajlarındanır [11].

Eklemeli sentezleyiciler konuřmayı sesbirim, difon, trifon, seslem vb. gibi nceden kaydedilen ses paralarını belirli sinyal iřleme teknikleriyle bir araya getirerek oluřturmaya alıřırlar [3]. Bu amala, ilk nce birleřtirilecek paraların veritabanı hazırlanmalıdır. Kayıt ařaması zahmetlidir, ama hedef ses paralarının kaydedilen konuřma paralarından ayrılması daha fazla zaman almaktadır. Eklemeli sentezlemede yer alan temel adımlar Őekil 2.8'de gsterilmektedir [3].



Şekil 2.8. Eklemeli sentezleme sistemlerinde bulunan temel adımlar

Konuşmayı oluşturacak ses parçalarının belirlenmesinin ardından, ikinci adım olarak sayısal sinyal işleme bloğuna geçilir. Ses parçaları birbirlerine fazları, frekansları ve enerjileri uyumlu olacak şekilde birleştirilmelidir [3]. Bunu sağlayan yöntemler; Örtüştürmeli Ekleme (OLA: Overlap Add) olarak bilinmektedir [3]. Yapılan işlem; genellikle Hanning Penceresi kullanılarak çerçevelerin çarpılması ve yumuşatılan bu bölümlerin %50 oranında örtüştürülmesi ile toplanarak birleştirilmesidir [3]. Bu yöntemin bilinen ve en çok kullanılan türleri PSOLA (Pitch Synchronous Overlap Add) ve WSOLA (Waveform Similarity Overlap Add) yöntemleridir [17,40-42].

Özetlenecek olursa MKS sistemlerinde sentezlenen konuşmanın doğallığı, üç temel unsura bağlıdır. Bu unsurlar konuşmada kullanılan sedalı seslerin perde frekansı, kullanılan fonemlerin enerjileri ve kullanılan fonemlerin süreleri olarak sıralanır [19].

## **2.6. Başlıca MKS Sistemleri**

Şuan gerek ticari gerekse akademik olarak geliştirilen birçok MKS sistemi bulunmaktadır. Bunlardan en çok bilinen MKS sistemleri Mitalk, Infovox, Bell Labs Tts System, Cnet Psola, Eti Elequence, Festival Tts System, Mbrola Ve Whistler'dir.

### **2.6.1. MI TALK**

1979 yılında Allen, Hunnicutt ve Klatt tarafından bir formant sentezleyici olarak hazırlanan sistem MIT laboratuvarlarında geliştirilmiştir [1]. Bugünkü birçok sistemin temelini bu sistemde kullanılan teknoloji oluşturmaktadır [1, 20].

### **2.6.2. INFOVOX**

Telia Promotor AB Infovox en popüler çok dilli MKS sistemlerinden biridir. İlk ticari sürümü 1982 yılında “İsveç Royal Teknoloji Enstitüsünde” geliştirilmiştir. Bu sistemde kullanılan yöntem azalan formant sentezidir[1].

Günümüzde halen bu sistemin yazılım ve donanım uygulamaları bulunmaktadır. Şuan bulunan Infovox230 da kadın, erkek ve çocuk seslerini içeren Amerikan ve İngiliz İngilizcesi, Danca, Fince, Fransızca, Almanca, İzlandaca, İtalyanca, Norveççe, İspanyolca, İsveççe ve Flamanca dilleri bulunmaktadır ve konuşma anlaşılabilir akıcılıkta olup sisteme yeni sesler de eklenebilmektedir [1, 21].

### **2.6.3. BELLABS TTS SYSTEM**

Difonların veya trifonların birleştirilmesine dayalı olan bu sistemin İngilizce, Almanca, Fransızca, İspanyolca, Rusça, Romence, Çince, Japonca sürümleri bulunup diğer diller için geliştirme çalışmaları devam etmektedir. İngilizce haricindeki diğer tüm dillerde sistem aynıdır. Bu yüzden çok dilli bir sistem olarak adlandırılabilir [1].

Dile özgü gerekli bilgiler ayrı tablo ve parametre dosyalarında saklanır. İyi bir metin analiz yeteneği, kelime veya uygun isim telaffuzu, tonlaması, parçalar arası süre, aksan ve cümle vurgulama sistemin başarıyla uyguladığı karakteristik özelliklerinden birkaçıdır [1].

Sistemin bir diğerk özelliđi modüler olmasıdır. Böylelikle farklı araştırma grupları farklı modüller üzerine birbirinden bağımsız çalışabilirler. Geliştirilen modüller istendiđi zaman sisteme eklenebilmektedir [1].

#### **2.6.4. CNET PSOLA**

Fransa Telecom CNET, 80'lerin ortasında birleştirilmiş sentez için en umut vaat eden metotlardan biri olan PSOLA'yı kullanan difon tabanlı sentezleyiciyi tanıtmıştır. En son ticari ürünü atasözü TTS sistemi olarak Elan Informatique sistemidir. Sistemde ton ve konuşma oranı ayarlanabilmektedir. Günümüzde Amerikan ve İngiliz İngilizcesi, Almanca, Fransızca ve İspanyolca dillerini bulunmaktadır [1].

#### **2.6.5. ETI ELEQUENCE**

Sistem Eloquent Teknoloji Inc. USA tarafından geliştirilmiştir. Günümüzde İngiliz ve Amerikan İngilizcesi Almanca, Fransızca, İtalyanca, Meksika ve Kastilya İspanyolcası dillerinde bulunmaktadır. Her dil için 7 ayrı ses vardır ve kullanıcı tarafından kolayca özelleştirilebilir [1,22].

#### **2.6.6. FESTIVAL TTS SYSTEM**

Sistem Edinburg Üniversitesinde CSTR'de geliştirilmiştir. İngiliz ve Amerikan İngilizcesi İspanyolca ve Gal dili desteklediđi dillerdir. Sistem eğitim, araştırma ve bireysel kullanım için ücretsiz temin edilebilmektedir [1].

#### **2.6.7. MBROLA**

Bu çalışma Belçika Mans Politeknik Fakültesindeki TCTS laboratuvarları tarafından başlatılan bir projedir. Projenin ana hedefi ticari olmayan ve araştırma odaklı kullanım için çok dilli bir MKS sistemi yaratmaktır. Bu projede kullanılan metot Psola' dakine çok benzerdir. Psola CNET'in markası olduđu için bu proje MBROLA olarak adlandırılır [1,23].

MBROLA tam bir TTS sistemi değildir. Çünkü ham metni girdi olarak kabul etmez. Fonemler ile duraksama ve ton gibi bazı prosodik bilgilerin listesini alır ve konuşmayı o şekilde üretir. Difon veritabanları Amerikan, İngiliz ve Bretonin gilizcesi, Portekizce, Fransızca, Flamanca, Almanca, Romence, İspanyolca ve Türkçe dillerinde bulunmaktadır. Tüm dil versiyonları için kadın ve erkek sesleri mevcuttur [1, 23].

### **2.6.8. WHISTLER**

Bu sistem Microsoft tarafından geliştirilmekte olan eğitilebilir konuşma sentezleme sistemidir. Sistemin amacı doğal sesli konuşma üretmek ve orijinal konuşmacının “akustik” ve “prosodik” karakteristiklerine benzeyen bir çıktı üretmektir. Konuşma motoru birleştirilmiş senteze ve “Hidden Markov” modellerindeki eğitim prosedürüne dayalıdır [1].

### **3. TÜRKÇE METİNDEN KONUŞMA SENTEZLEME SİSTEMLERİ**

Türkçe, Ural-Altay dilleri grubuna bağlı eklemeli bir dildir bu yüzden Türkçe metinler için konuşma sentezlemede en uygun yöntemin Eklemeli Sentezleme olduğu belirtilmektedir [4].

#### **3.1. Türkçe Dil Ailesi ve Özellikleri**

Türkçe, diğer Türk dilleriyle birlikte Altay dil ailesinin bir kolunu oluşturmaktadır. Bu ailenin diğer üyeleri Moğolca, Mançu-Tunguzca ve Korece'dir. Türkçe, diğer Altay dilleri gibi eklemeli, yani sözcüklerin eklerle yapıldığı ve çekildiği, sondan eklemeli bir dildir.

Dilde özleşme çabaları 19. yüzyılın ikinci yarısında Tanzimat dönemi ile başlayıp aydınların Türkçe sözcük kullanma ve Arap alfabesinde yenilik çabalarıyla geçen bir hazırlık döneminden sonra Cumhuriyetle birlikte çağdaş Türkçenin temelleri atılmıştır [24]. Atatürk'ün özel ilgi ve çabalarıyla Latin alfabesine geçilmiş; tarama, derleme ve türetme yoluyla dildeki Türkçe sözcük oranı kısa sürede büyük oranlara ulaşmıştır. Türkçe dili morfolojik olarak eklemeli bir dil olduğundan, bir sözcükten onlarca sözcük türetilmektedir [25]. Bu sebeple dildeki sözcük sayısı kolaylıkla artırılabilir.

#### **3.2. Ses, Fon, Fonem ve Hece Kavramları**

Bir ses sinyali genellikle ünlüler ve ünsüzler olmak üzere ikiye ayrılmaktadır. Bu terimler sesbiliminde kullanılmasına rağmen, bunları kesin olarak tanımlamak zordur. Ünlü ses sinyalleri tek başına seslendirilebilirken ünsüz ifadeler seslendirmede mutlaka bir ünlü ifadeye ihtiyaç duyulmaktadır [6].

### 3.2.1. Ses Fonem Ayrımı

Sesbilim (phonetics), herhangi bir dilden bağımsız olarak konuşma sinyallerinin bir görüntüsü iken, fonem bilim (phonemics) belirli bir dile ait konuşma sinyallerinin bir görüntüsüdür. Sesbilim, dilbilimine kaynak sağlarken; fonem bilim ise dilbiliminin bir kolu olmaktadır. Sesbilimde, en küçük birim ses (phone) olarak adlandırılırken, fonem bilimde en küçük birim fonem (phoneme)'dir. Sesler köşeli parantezler içinde “[b]”, fonemler ise, eğik çizgiler arasında “/b/” yazılırlar [6].

Fonemin bir tanımını yapacak olursak belli bir dilde bir kelimedenden diğerine fark eden en küçük ses birimi olarak tanımlamak mümkündür. Eğer herhangi bir sesli ifadede bir ses biriminin değiştirilmesi o sesli ifadenin anlamını da değiştiriyor ise, bu durumda değiştirilen birim aynı zamanda fonemdir. Eğer o ses biriminin değiştirilmesi herhangi bir farklılık yaratmıyor, ya da yeni oluşan sesli ifade o dil bağlamında anlamsız oluyor ise bu durumda değiştirilen birim bir fonem değildir [6].

Seslerin sayısı, onları ayırt etme yeteneğimizle sınırlı olup aslında çok geniştir, fonemlerin sayısı ise bütün dillerde küçüktür. Bilinen en büyük fonem sayısı 45 ile Chipewyan diline, en küçük ise 13 fonemle Havai diline aittir [6].

Dilin en yalın ögesi sestir. Sesli ifadeler yazıya geçirildiğinde her sese karşılık bir alfabetik simge gelmektedir. Yalnız ses sayısı ne kadar fazla ise sesli ifadeden yazıya geçiş o kadar karmaşıktır. Bu nedenle kümelenme yoluna gidilmektedir. Bu yöntemde kimi sesler kümelenecek her küme için tek bir simge verme yolu kullanılmaktadır. Bu bağlamda seslerin kümeleneceği fonem tabanında gerçekleşir. Fonemler, anlam ayırıcı özelliği bulunan ses kümeleridir. Yani anlam ayırıcı özelliği bulunmayan sesler tek bir fonem kümesi altında toplanır. Ses, diller üstü bir birimdir. Fonem ise anlam ayırıcı özellik içermesi nedeniyle belli bir dile özgü birimdir [6, 26].

Her dilin kendine özgü bir alfabesi bulunmaktadır ve bu bağlamda Türk Alfabesi incelendiğinde fonemik bir alfabe olarak söylenebilir. Türkçenin her fonemine bir alfabetik simge (harf) atandığı bilinmektedir. Sesler bir fonem kümesine dahil edilirken izlenen yol, anlam ayırımına dayanmaktadır yani eğer bir sözcük içerisinde yer alan bir ses, başka bir sesle yer değiştirdiğinde bir anlam farklılığı oluyor ise, bu sesler farklı fonemlere aittir [6]. Örneğin “not” ile “net” örneğinde değiştirilen ses ile sözcüğün anlamı da değiştiği için, “o” ve “e” seslerinin farklı fonem kümelerinde yer aldıkları söylenebilir.



Ses ise fonemlerin seslendirilmesiyle oluşur [6]. Diğer bir açıdan seslerin yazıya dökülerek ifade edilmesi sonucu oluşan simgeler fonem olarak adlandırılır.

Fonemlerin tanımlanabilmesi için ayırıcı özellikler (distinctive features) kuramı geliştirilmiştir [6,27]. Fonemlerin ayırıcı özellikleri toplamda 12 adettir (Tablo 3.1). Her özellik birbirinden bağımsız ve olası iki değerden yalnız bir tanesini alabilmektedir [6].

Tablo 3.1. Fonem ayırıcı özellikleri [6].

1.Sessellik (Vocalic/Nonvocalic):	Tanımlı bir formant yapısının var olup olmamasına karşılık gelir.
2.Ünlü/Ünsüz (Consonantal/Nonconsonantal):	Fonemin ünlü veya ünsüz olmasıyla ilgilidir. Ünlü özelliğindeki fonemler görece olarak daha küçük bir toplam enerjiye sahiptir.
3.Sıkışık/Seyrek (Compact/Diffuse):	İzgesel (spectral) enerjinin dağılımına karşılık gelir.
4.Gergin/Gevşek (Tense/Lax):	Gergin, daha geniş bir bant genişliği ve uzun bir sürede daha yüksek bir toplam enerjiye karşılık gelir.
5.Ötümlü/Ötümsüz (Voiced/Voiceless):	Ses tellerindeki titreşimlere bağlı olarak düşük frekanstaki bileşenlerin varlığına dair bir özelliktir. Ötümlü bir fonemde bu bileşenler bulunur.
6.Genizden/Ağızdan (Nasal/Oral):	Genizden meydana getirilen fonemler fazladan rezonansların eklenmesi ile daha geniş bir izgesel enerji dağılımı gösterir.
7.Kesilmeli/Sürekli (Discontinuous/Continuous):	Kesilmeli fonemler izgesel enerji yayılımında ani enerji değişiklikleri gösterir.
8.Keskin/Yumuşak (Strident/Mellow):	Keskin fonemler daha güçlü ve daha fazla gürültü bileşenleri içerir.
9.Patlamalı/Patlamasız (Checked/Unchecked):	Patlamalı fonemlerde enerji, patlama şeklinde görülür.
10.Bas/Tiz (Grave/Acute):	Bas seslerde düşük frekanslı rezonanslar hakim iken, tiz seslerde yüksek frekanslı rezonanslar hakimdir.
11.Flat/Plain:	Yüksek frekanslı rezonansların görece enerjilerine göre farklılık göstermektedir. Flat daha zayıf, plain ise daha güçlü özellikteki fonemleri tanımlar.
12. Keskin/Düz (Sharp/Plain):	Keskin fonemler, daha yüksek frekanslı rezonansların görece frekanslarında bir yükselme gösterirler.

Sesbirim kavramı genişletilerek parçalı sesbirimler ve parçalar üstü sesbirimler olarak ikiye ayrılır [6, 28]. Parçalı sesbirimleri, ünlü, ünsüz ve kayan ünlüleri içerir. Bu sesbirimler saptanırken anlam ayırıcı özelliklerinden faydalanılmaktadır. Bunun dışında süre, perde değişimi, ton, vurgu, ezgi gibi ayırıcılardan da faydalanılması ile meydana gelen sesbirimlere, parçalar üstü sesbirimler adı verilir ve parçalar üstü sesbirimleri sesli ifadenin bir kesimine bakarak belirlenemez, genellikle tamamına bağımlıdırlar [6].

### **3.2.2. Hece Kavramı**

Konuşma, birim konuşma seslerinin ardı ardına seslendirilmesi ile meydana gelir. Fonemler ise birleşerek heceleri meydana getirir. Bu kural bütün dillerde geçerlidir [6]. Ancak heceler farklı dillerde farklı biçimlerde olabildiği için heceleri tanımlamak zor bir süreçtir. Her dil için heceler, yapıları bakımından farklılık göstermektedir. Bu yüzden tüm dilleri kapsayacak ortak bir hece tanımı yapmak mümkün olmamaktadır.

Türkçe dili içindeki her hecede mutlaka bir sesli harf bulunur, sesli harf olmadan hece kurulamaz. Sesli ifadeler, kaburgalar arası kaslarca biçimlendirilen göğüs atışları neticesinde oluşurlar. Her göğüs atışıyla birlikte ses telleri de titreşmeye başlar. Bu şekilde ünlü sesin çıkarılması gerçekleşir [6]. Tüm hecelerde tek bir sesli harf olabileceği için, her göğüs atışı sürecinde çıkarılan sesler hece olarak da söylenebilir. Bu süreç boyunca belirli basınçta soluk, ses yolundan dışarı çıkar. Bu basınçlı havaya ses tellerinin titreşimlerinin eklenmesiyle ünlü sesler çıkarılır [6]. Kasların gevşemesiyle birlikte geçen havanın azalması ya da durdurulması sırasında ses tellerinin titreşmesi ya da durması ünsüz seslerin çıkarılmalarını sağlar [6,26].

Heceler, kelimelerin ses yapısını oluştururlar. Türkçede sesli harfler tek başlarına hece özelliği gösterdikleri halde sessiz harfler yanlarına ünlü almadan bir hece oluşturamazlar. Dolayısıyla Türkçe bir sözcükte kaç sesli harf varsa, o kadar da hece var demektir. Çünkü Türkçe bir hecede birden fazla sesli harfin bulunması mümkün değildir. Sessiz harfler, kendilerini takip eden sesli harflerle birleşerek hece oluştururlar.

### 3.3.Türkçe 'de Harf Ses İlişkisi

Türk alfabesinin fonetik bir alfabe olduğu yani harf-ses dönüşümünün bire bir (grapheme - to - phoneme mapping) olduğu söylenmektedir [29]. Ancak bu durumun bozulduğu çok sayıda örnekte bulunmaktadır. Bu tip durumlar özellikle yabancı (Fransızca, Arapça, Farsça) kökenli sözcüklerde sıkça görülmektedir. Türkiye Türkçesindeki seslerin alfabetik gösterimi hakkında gerek dilbilimcilerin [29,30], gerekse konuşma tanıma ve sentezleme ile ilgilenen araştırmacıların [31] çeşitli çalışmaları bulunmaktadır. Bu çalışmalardaki ortak çıkarım, 29 harfin günümüz Türkçesinde konuşma esnasında çıkartılan seslerin tamamı ile temsil edilmesinde yetersiz kaldığıdır [5]. Örneğin bir dilbilimci olan Ergenç'in çalışmasında ise [16], günümüz Türkçesinde 44 farklı ses olduğu belirtilmiştir.

1928 yılında Harf Devrimi yapılırken “^” düzeltme işaretinin hem uzatma, hem de inceltme amacıyla kullanımının yeterli olacağı öngörülmüştür [32]. Dolayısıyla bu işaret, bazen üzerinde bulunduğu sesli harfin uzatılması, bazen de üzerinde bulunduğu sesli harfin önündeki sessiz harfin (örneğin k, g ve l) inceltmesi amacıyla kullanılmaktadır. Ancak tek bir düzeltici işaretin yetersiz kaldığı durumlar da bulunmaktadır. Örneğin “garip, gazi, gâvur,cengâver” sözcükleri ele alındığında; ilkinde “ga” hecesinin normal okunduğu, ikincisinde “a” harfinin uzun okunduğu, üçüncüsünde “g” harfinin ince okunduğu,sonuncusunda ise “g” harfinin ince okunmasının yanı sıra “a” harfinin de uzun okunduğu görülmektedir [5].

Son 20–25 yıl içinde, gerek Türk Dil Kurumu'nun mevcut düzeltme işaretini (^) kaldırdığı yönündeki asılsız söylentiler [32], gerekse bilgisayar, cep telefonu gibi sadece standart yazı karakterlerini destekleyen yazım ortamlarının sıkça kullanılmaya başlanması sebebiyle Türk alfabesinin zaten yetersiz olan tek düzeltici işareti de uygulamada iyice az kullanılır hale gelmiştir. Her şeye rağmen düzeltici işaret eksikliği,sözcükleri bir bütün halinde görerek tümünden gelim tekniğiyle okuma yapan okuyucular için büyük bir telaffuz sorunu teşkil etmemektedir. Ancak yazılım destekli metinden konuşma sentezleme uygulamalarında, doğru telaffuz için gerekli yerlerde ilgili düzeltici işaretlerin kullanımı, büyük önem arz etmektedir [5].

Harf-ses dönüşümünde bire bir olma özelliğini bozmakta olan istisnai 154 hece tespit edilmiştir [2]. Söz konusu heceler ve bunların farklı okunuşlarına dair örnekler, Tablo 3.2'de verilmiştir.

Tablo 3.2. Okunuşları ve yazılışları bire bir olmayan istisnai heceler [2].

Hece	Örnekler	Hece	Örnekler	Hece	Örnekler
a	atik, âcil	fi	filiz, terfî	la	lala, aslâ, lâma, lâle
al	alkış, âlköl	fol	folluk, fölyo	laç	kulaç, ilaç
at	at, itaât	ga	garip, gâzi, gâvur, cengâver	lah	allah, ilâh
ba	barış, bâriz	gal	mangal, işgâl	lak	parlak, aylâk
bal	bal, istikbâl	gan	slogan, bezirgân	lam	sağlam, lâmba
bel	belli, bëlge	gar	gar, rüzgâr	lan	yalan, rölânti
bem	bembeyaz, şübem	gat	ırgat, ferâgat	lar	toplardamar, bilârdo
ben	benzin, bënlik	gul	murgul, meşgûl	las	atlas, iflâs
ber	berrak, habër	ha	hayır, hâki	laş	buharlaşma, telâş
bi	birâ, bîtap	hal	halhal, derhâl	lat	vuslat, cellât
bol	futbol, karambòl	hat	rahat, sıhhât	lay	olay, kelâynak
bu	budak, bûse	hem	hemcins, dirhëm	li	renkli, millî
bul	bulgur, kabûl	hi	hile, hîbe	lo	kablo, kolóni
ca	damacana, ticâri	hu	huni, hûri	lom	slalom, lómboz
cel	cellat, güncël	i	inek, îkaz	lon	kolonya, teflón
cem	cemre, düşüncëm	ja	pijama, jâle	lu	oluşum, tulúat, ulýfe
cen	cendere, sevecën	ka	karışık, kâbiliyet, kâğıt, kâtip	luk	oluk, mahlúk
cer	cerrah, hacër	kah	kahve, kâhya	lum	tulum, mâlúm
ci	eskici, fecî	kal	kalkan, kâlker	lun	pulun, melún
cu	burcu, ucûbe	kam	kambur, ahkâm	lup	olup, üslúp
ça	çarık, çâre	kan	kan, dükkân	lur	olur, billúr
çem	çember, perçëm	kar	kartal, efkâr	lut	bulut, lút
da	damızlık, dâhil	kat	sakat, dikkât, zekât	ma	makara, mâvi
dal	pedal, madâlya	kel	kelle, kël	mal	mal, ihmâl
dem	demlik, madëm	kem	kemre, hakëm	me	mezura, mêmur
den	denge, madën	ken	kendir, erkën	mem	membâ, gitmëm
di	diyar, ciddî	ki	kitap, hakîki	men	mendil, göçmën
dol	dolgu, idòl	kol	karakol, alköl	mi	mine, resmî
du	durum, gudûbet	ku	kuma, sükýnet	mo	moruk, limônî
fa	kafa, fâre	kul	kul, mâkûl	mu	murat, umûmî
fal	fal, kefâl	kum	kum, mahkúm	mü	müsait, mýmin
fel	fellik, fël	kun	suskun, meskún	na	nasır, nâne, manâ
fen	fennî, lütfën	kut	yâkut, sükút	nal	sanal, banâl
nem	nem, önem	sem	sembòl, sersëm	tü	tüzel, temettý
ni	nişan, nîsan	sen	senkron, sën	u	ucuz, ûdî
nu	konu, nûri	si	sini, kasîde	ul	ultraviyole, ùlvî
pa	para, pâyeye	sol	sollamak, turnusòl	va	vana, vâkî
pal	palto, pâlmiye	su	sulu, sûdan	val	karnaval, festivâl
pen	pencere, lümpën	sul	sultan, mesûl	vi	vize, âvîze
pi	piyaz, pîrî	şa	şaka, şâir	vu	vuruş, davûdî

Tablo 3.2.'nin Devamı: Okunuşları ve yazılışları bire bir olmayan istisnai heceler [2]

Hece	Örnekler	Hece	Örnekler	Hece	Örnekler
pol	polka, sivastopòl	şal	şal, şalter	ya	oya, rüyâ
pul	pul, ampül	şu	şurup, şûrâ	yal	yalnız, hayâl
ra	rakı, râzı	ta	tazı, tâyin, hattâ	yel	yellenmek, yël
ral	kural, râlli	tal	hantal, iptâl	yem	yemyeşil, gayëm
rat	surat, sürât	te	temiz, têsis	yen	yenge, duayën
rem	remzi, harëm	tel	tellâk, tëlgraf	za	zafer, imzâ
ren	kahverengi, mürën	tem	temsil, mâtëm	zal	azalmak, inzâl
ri	huri, ârî	ten	tenzil, satën	zem	zemheri, elzëm
rol	erol, röl	ti	tiraj, âtî	zen	zengin, bâzën
ru	rubâî, rûhi	tol	tolga, atòl	zi	arâzi, farazî
sa	sapık, sâde	tu	tutum, rutûbet	zu	kuzu, fuzûlî
sal	yasal, misâl				

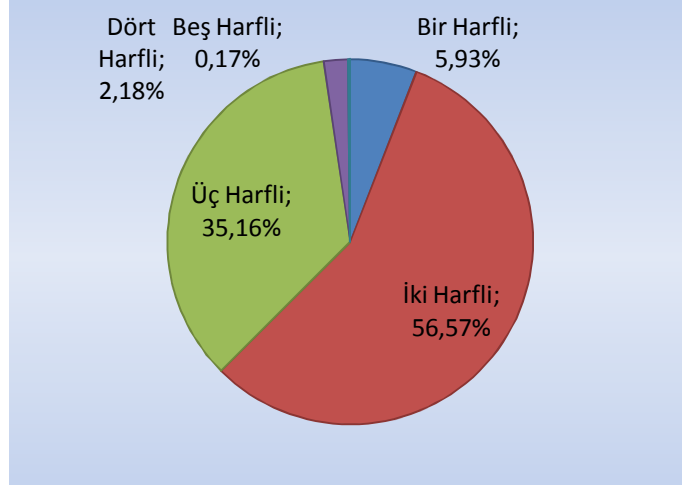
### 3.4. Türkçede Bulunan Hece Türleri

Türkçe dilinde heceler en az bir, en çok dört harften oluşur ve bazı özel heceler “bal, kol, dal, çal, kürk” dışında diğer heceler anlamsızdır. Tablo 3.3’te “C” sessiz ve “V” sesli harfleri belirtmek üzere Türkçe hecelerinin genel yapısı verilmiştir [4].

Tablo 3.3. Türkçede hecelerinin genel yapısı [4].

Hece Yapısı	Örnek heceler
V	a, e, ı, i, o, ö, u, ü
VC	ab, ac, aç, ad, ... ,az, eb, ec, ...
CV	ba, be, bı, bi, ... , za, ze, zı, zi, ...
CVC	bel, gel, köy, tır, ...
VCC	alt, üst, ırk, ...
CCV	Bre
CVCC	kurt, yurt, renk, Türk

Şekil 3.1’de ise Türkçe hecelerinin uzunluklarının yüzde oranları verilmiştir. Görüldüğü gibi, Türkçe metinlerde en fazla iki harfli heceler yer almaktadır [4].



Şekil 3.1. Hece uzunluk yüzdeleri [4].

### 3.5. Yapılan Akademik Çalışmalar

Türkçe metinlerden konuşma sentezleme ve ilgili konularda, bugüne kadar yapılmış olan lisansüstü ve doktora çalışmaları Tablo 3.4'te verilmiştir [33]. Bu çalışmalar incelendiğinde, büyük çoğunluğunun eklemeli sentezleyiciler üzerine olduğu görülmüştür. Eklemeli Sentezleme Sistemlerinde birleştirilen parçaların önceleri sesbirim, daha sonra hece ve son çalışmalarda çoğunlukla difonlar olduğu görülmektedir. Ayrıca, eklemeli sentezleyicilerde, farklı sinyal işleme (birleştirme) yöntemlerinin (değişik örtüşüp ekleme yöntemleri, sinüzoidal model gibi) başarısı incelenmiştir. Son yapılan çalışmaların büyük bir bölümünde ise, daha doğal bir konuşma sentezi için, süre ve ezgi modelleri geliştirilmeye çalışılmıştır [33].

Tablo 3.4. Türkçe metinden konuşma sentezleme üzerine yapılan akademik çalışmalar [33].

Yazar	Çalışmanın Türü, Yılı	Üniversite	Çalışmanın Başlığı
İlhan Yaşar ÖZÜM	Yüksek Lisans, 1993	ODTÜ	“A Speech Synthesis System for Turkish Language Ased on the Concatenation of Phonemes Taken From a Speaker
Kamil GÜVEN	Yüksek Lisans, 1994	Çukurova	“PC Based Speech Synthesis for Turkish”
Murat Servet ERER	Yüksek Lisans, 1994	İTÜ	“Karma Söz Üretme Yöntemi ile Türkçe Yazılı Metinden Söz Etme”
Kerem AYHAN	Yüksek Lisans, 1998	ODTÜ	“Text to Speech Synthesis Turkish Using Non Parametric Techniques”
Özgür SALOR	Yüksek Lisans, 1999	ODTÜ	“Signal Processing Aspects of Text to Speech Synthesis Turkish”
Barış BOZKURT	Yüksek Lisans, 2000	Boğaziçi	“Reading Aid For Visually Impaired (A Turkish Text-To Speech System Development)”
Çağla ÖNÜR	Yüksek Lisans, 2001	ODTÜ	“Concatenative ve Speech Synthesis Based on a Sinusoidal Speech Model “
Erkan ABDULLAHBEŞE	Yüksek Lisans, 2001	Boğaziçi	“Fundamental Frequency Contour Synthesis for Turkish Text to Speech“
Şifa Serdar ÖZER	Yüksek Lisans, 2002	Hacetepe	“Türkçe Metinden Konuşma Sentezleme”
Barış EKER	Yüksek Lisans, 2002	Bilkent	“Turkish Text to Speech System”
Ömer ŞAYLI	Yüksek Lisans, 2002	Boğaziçi	“Duration Analysis and Modelling for Turkish Text to Speech Synthesis”
Banu Oskay	Yüksek Lisans, 2002	ODTÜ	Automatic Modelling of Turkish Prosody
Esra VURAL	Yüksek Lisans, 2003	Sabancı	“A Prosodic Turkish Text to Speech Synthesizer“
Oytun TÜRK	Yüksek Lisans, 2003	Boğaziçi	“New methods for Voice Conversion
Haşim SAK	Yüksek Lisans, 2004	Boğaziçi	“A Corpus Based Concatenative Speech Synthesis Sytse for Turkish”
Ozan AKTAN	Yüksek Lisans, 2004	Boğaziçi	A Single Chip Solution text to Speech Synthesis”
Asude KARLI	Yüksek Lisans, 2005	Ankara	“Örnek Bir Dizi Cümle İçin Türkçe Metinden Konuşma Sentezleyici”
Ögül SALOR	Doktora, 2005	ODTÜ	“Voice Transformation and Development of Related Speech Analysis Tools for Turkish”
Özlem ÖZTÜRK	Doktora, 2005	ODTÜ	“Modelling Phoneme duration and Fundamental Frequency Contours in Turkish Speech
İlker ÜNALDI	Yüksek Lisans, 2007	Hacetepe	“Taşınabilir Cihazlar İçin Türkçe Metinden Konuşma Sentezleme Sistemi”
Oytun TÜRK	Doktora, 2007	Boğaziçi	“Cross Lingual Voice Conversion

## **4. TÜRKÇE METİNLER İÇİN EKLEMELİ SENTEZLEME YÖNTEMİ**

### **4.1. Türkçe Metinden Konuşma Sentezleme Sistemi Yapısı**

Türkçe MKS sistemleri 4 temel basamaktan oluşmaktadır. Bunlar; metin ön işlemlerinin yapılması, Metnin Hecelere Ayrılması, Ses Veri tabanı Oluşturulması ve seslerin birleştirilmesi işlemidir.

### **4.2. Metin Normalizasyonu**

Metin ön işleme aşamasında rakamlar, sayılar, kesirler, tarihler, sıra belirten ifadeler, kısaltmalar ve özel karakterler gibi yazı dilinde anlamı olan ifadeler, okunurken sarf edilen sözcüklere dönüştürülmelidir. Buna “standartlaştırma veya normalizasyon” da denebilir.

#### **4.2.1. Kısaltmalar**

Kısaltmaların okunması da dikkat isteyen bir iştir. Kısaltmalar, açılımlarına dönüştürülebilir, harf harf okunabilir, kısaltmanın kendisi bir kelimeymiş gibi okunabilir veya kısaltma bambaşka bir şekilde okunabilir [15].

İlk olarak kısaltma ve kelimeleştirilmiş kısaltmaların tespiti yapılmalı, bu istisnai ifadelerin tespiti için kısaltmalar sözlüğü oluşturulmalıdır. Büyük harf ile yazılan kısaltmalar ve kelimeleştirilmiş kısaltmalar (TSK, TDK, TUBİTAK vb.), daha sonra “nokta (.)” içeren kısaltmalar (“Dr.”, “Doç.”, “Prof.” vb.), son olarak da küçük harf ile yazılan ve sonunda “nokta (.)” yer almayan kısaltmalar (mm, kg, l, K, vb.) tespit edilmeli bunların ayrı bir veritabanından çağırılması gerekmektedir [25].

Yazım kurallarına göre küçük harfli kısaltmalar, kısaltmanın uzun haline göre, bir başka deyişle ilgili sözcüğün okunuşuna göre ek alırlar.



Örneğin “kg.’ına” kısaltması kilogramına, “ml.’sinde” mililitresinde şeklinde seslendirilmektedir. Dolayısıyla metinden konuşma sentezleme uygulamalarında, bu tarz kısaltmaların metin içerisinde uzun hallerinin seslendirilmesi daha uygun olacaktır [2].

Yazım kurallarına göre büyük harfli kısaltmalar ise kısaltmanın okunuşuna göre ek alırlar. Örneğin TBMM’de “tebememedede”, AB’nin “abenin”, ODTÜ’ye “odtüye” şeklinde okunmaktadır. Dolayısıyla metinden konuşma sentezleme uygulamalarında, bu tarz kısaltmaların metin içerisinde kısaltma olarak seslendirilmesi daha uygun olacaktır. Büyük harfli kısaltmaların okunuşu hakkında şu tespitleri yapmak mümkündür [2].

TBMM, TPAO gibi kısaltmalar her bir harfin alfabetik seslendirilmesi yapılarak; bir başka deyişle sesli harfler uzatılarak, sessiz harflerin sonuna ise uzun “e” getirilerek okunmaktadır. Ancak PTT ve TRT gibi bazı kısaltmalar hızlı yani sessiz harflerin sonuna normal “e” getirilerek okunmaktadır [2].

CD (“sidi”), DVD (“dividi”) gibi bazı kısaltmalar ise orijinal dillerindeki gibi okunmaktadır. Ayrıca “k” harfi bazı kısaltmalarda “kâ”, bazılarında ise “kê” şeklinde okunmaktadır. Örneğin SPK “sepeka” ve MKE “mekee” şeklinde okunmaktadır [2].

“NATO”, “ASELSAN”, “TÜBİTAK” gibi kelimeleşmiş olan kısaltmalar ise aynen okunmaktadır. Öte yandan yine kelimeleşmiş olan “UNESCO”, “UNICEF”, “OPEC” gibi yabancı kökenli kısaltmalar ise ilgili dildeki telaffuz kurallarına göre okunmaktadır [2].

#### **4.2.2. Sayılar**

Sayı değerlerinin seslendirilmesi için yazılı hale getirilmesi, metinden konuşma sentezleme uygulamalarında önem arz eden bir diğer işlemdir. Tamsayı, ondalıklı sayı, tarih, saat, telefon, TC Kimlik Numarası gibi özel bilgileri belirten sayısal değerlerin her biri, farklı bir şekilde okunmaktadır. Örneğin: “1453” sayısı “bindörtüüzelliüç” şeklinde okunacak şekilde dönüştürülmelidir. Bu şekilde rakamları yazıya dönüştürüp bu şekilde seslendirilmeleri yapılmalıdır. Benzer biçimde, “3/7” kesir ifadesi “3 bölü 7” veya “yedide üç” şeklinde, “10.10.2010” veya “10/10/2010” gibi tarih ifadeleri de “on ekim ikibinon” veya “on on ikibinon” şeklinde çözümlenebilmelidir [2].

### 4.2.3. Özel İşaretler

Kısaltma ve sayıların tespitinden sonra metin içerisindeki noktalama işaretleri tespit edilmeli, “virgül (,)” için yarım ölçü ve “nokta (.)” için bir tam ölçü sessizlik eklenmektedir. Bunların dışında metin içerisinde geçen “+, -, /, \$, %” vb. işaretlerin doğru bir şekilde okunması gerekmektedir. Örneğin “+” işareti bir sayının önüne geldiğinde “artı” olarak okunup; kan grubu gibi özel bir değer belirtilmesi durumunda “pozitif” olarak okunabilmelidir [2].

### 4.2.4. Dilin Seslendirmedeki Farklılıkları

Türkçe’de yazı dili ile konuşma dili arasında bazı farklılıklar bulunmaktadır. Ulama, sessiz yumuşaması, ses düşmesi bu durumlara örnek olarak verilebilir.

Ulama sessizle biten bir kelimedenden sonra sesliyle başlayan bir kelime gelmesi durumunda sessiz harfin yumuşayarak diğer kelimenin sesli harfi ile birlikte okunması durumudur. Örneğin, “tat + almak” “ta-dal-mak”, “talep + etmek” “ta-le-bet-mek”; “öç + almak” “ö-calmak”. Ulama dilimizde sıkça kullanılan seslendirme kurallarından birisidir.

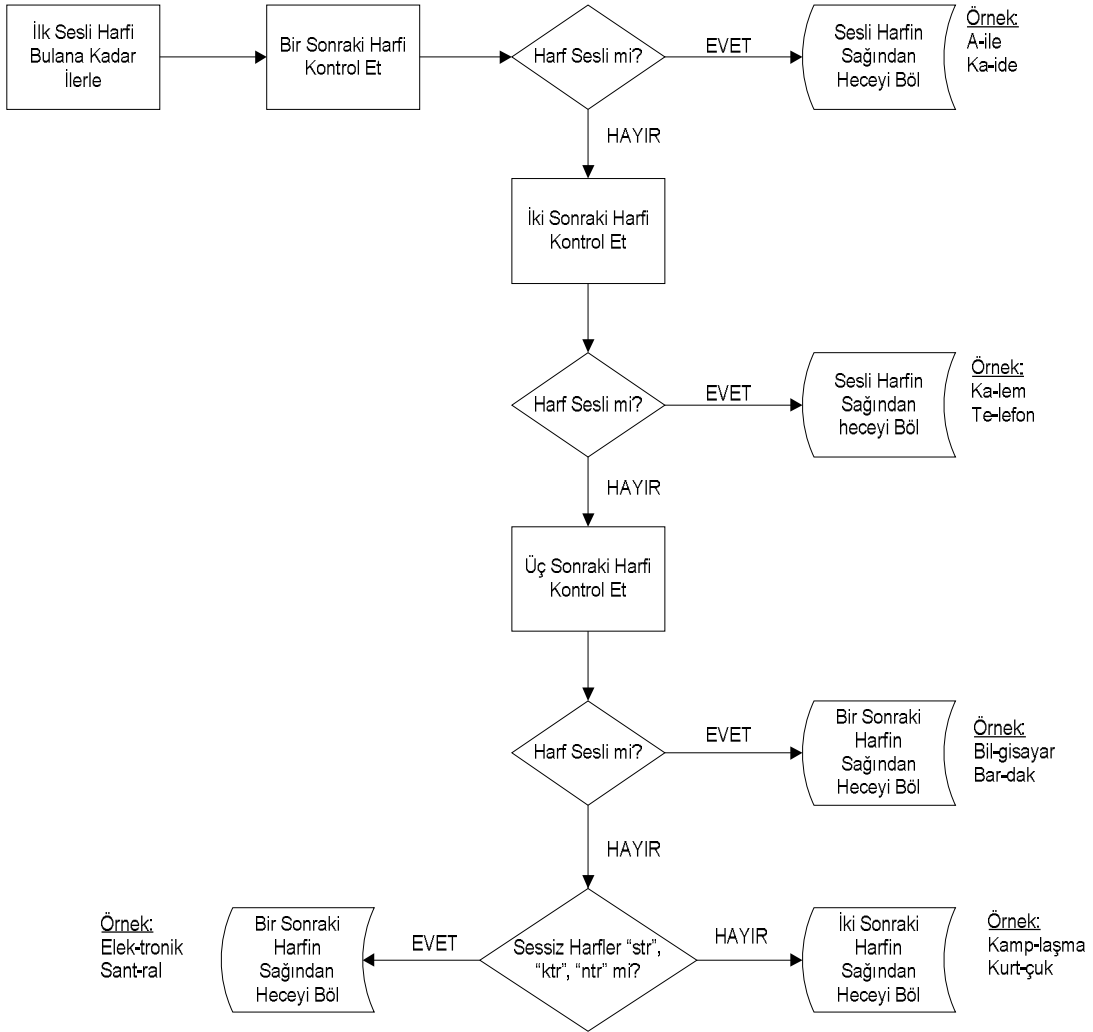
Türkçe’de bir isim “etmek”, “olmak”, “eylemek” gibi fiillerle birlikte kullanıldığında ise ses düşmesi meydana gelebilmektedir. Örneğin, hapis + etmek “hapsetmek” gibi. Öte yandan isim ile eylem arasında kaynaşma olması durumunda ise ses eklenerek okunabilmektedir. Örneğin, “zan” + “etmek” “zannetmek” gibi [2].

Sert sessiz yumuşaması, özel isimlerde yazım esnasında uygulanmamakta, okuma esnasında uygulanmaktadır. Örneğin “Irak’a” olarak yazılan sözcük, “İrağa” olarak; “Eyüp’ü” olarak yazılan sözcük ise “Eyübü” olarak okunmaktadır.

Bunların dışında Dilbilimsel çözümleme aşamasında sözcüğün cümle içindeki anlamına göre seslendirme yoluna gidilmelidir. Örneğin “Fatma hala gitmedi” cümlesinde bulunan “hala” sözcüğü iki anlamda kullanılabilir. Babanın kız kardeşi anlamında kullanılmışsa sert okunması gerekirken, henüz anlamında kullanılmışsa yumuşak seslendirilmelidir. Ancak bu cümlede hangi anlamda kullanıldığının tespiti konusunda kesin bir yöntem olmadığı için uygulama esnasında kullanılmamıştır.

### 4.3. Metnin Hecelere Ayrılması

Heceleme algoritması, temelde literatürdeki çeşitli kaynaklarda [4,34] verilmiş olan Türkçe sözcüklerdeki sesli-sessiz harf dizilimine göre geliştirilmiştir. Ancak, yabancı kökenli, özellikle Latince köklerden türetilmiş bileşik sözcüklerde yanlış heceleme engellemenek için söz konusu algoritmaya özel bir takım kurallar eklenmiştir [2]. Algoritmanın akış diyagramı, Şekil 4.1'de görülmektedir.



Şekil 4.1. Heceleme algoritması [2].

#### 4.4. Sesler

Ses, gırtlaktaki ses tellerinin hava moleküllerini titreştirmesi sonucunda, bu hava moleküllerinin sıkışma ve dağılması ile enerjinin uygulandığı yöne paralel olarak boyuna basınç dalgaları oluşturması sonucunda meydana gelir [35].

“Nyquist” teoremine göre kayıpsız bir sayısallaştırma için örnekleme oranının frekans yanıtının iki katı ve üstünde bir değer olması gerekmektedir. İnsanın işitsel frekans aralığı 20 Hz – 20 kHz olduğuna göre en fazla 40kHz örnekleme yeterli olmaktadır. Konuşma frekans aralığı ise 200 Hz – 8 kHz aralığında olursa en fazla 16 kHz örnekleme yeterli olacaktır [11,35].

##### 4.4.1. Temel Bilgiler

Bu tür çalışmalarda yapılan kayıtlar incelendiğinde genelde “wav” dosya formatının kullanıldığı görülmüştür [36]. Wav dosya formatı İngilizce’de “dalga” anlamına gelen “Wave” kelimesinin ilk üç harfinin alınmasıyla oluşturulmuş bir dosya formatıdır [35]. Sesin hiçbir sıkıştırma yöntemi uygulamadan saklandığı bir dosya biçimidir. IBM ve Microsoft’un küçük ses kayıtlarını herhangi bir bilgisayarda çalmak için ortaya çıkardığı bir dosya biçimidir [37]. Yapısı çok basittir. MP3 ve diğer sıkıştırılmış biçimlerin aksine Wav’lar sadece sayısallaştırılmış seslerdir. Basittirler, sıkıştırma uygulamamasının etkisiyle sesin kalitesini düşürmezler, ancak çok yer kaplarlar [35].

Wav dosyasında örnekler ham veri şeklinde sıkıştırılmadan tutulur. Wav dosyası üç veri bölgesi (chunk) içermektedir. RIFF veri bölgesi 12 byte uzunluğundadır ve dosyanın bir “wav” dosyası olduğunun belirtildiği bölgedir [36].

Tablo 4.1. Riff veri bölgesi [11].

Byte Sırası	Açıklama
0-3	“RIFF” (ASCII karakterleri şeklinde)
4-7	LittleEndian şekilde paketin geri kalan boyutu
8-11	“WAVE” (ASCII karakterleri şeklinde)

RIFF veri bölgesi alanları Tablo 4.1’de gösterilmiştir [11,35]. Sonraki veri bölgesi FORMAT veri bölgesidir. Bu bölgede formata özgü parametreler tanımlanmaktadır ve 24 byte uzunluğundadır. FORMAT veri bölgesi alanları Tablo 4.2’de gösterilmiştir [11].

Tablo 4.2. Format veri bölgesi [11].

Byte Sırası	Açıklama
0-3	“RIFF” (ASCII karakterleri şeklinde)
4-7	LittleEndian şekilde paketin geri kalan boyutu
8-9	“WAVE” (ASCII karakterleri şeklinde)
10-11	Kanal sayısı (0x01=Mono, 0x02=Stereo)
12-15	Hz olarak örnekleme oranı (binary)
16-19	Saniyedeki sekizli miktarı
20-21	Örnekteki sekizli miktarı 1=8 bit Mono, 2=8 bit Stereo yada 3=16 bit Mono, 4=16 bit Stereo
22-23	Örnekteki bit sayısı

Üçüncü veri bölgesi ise DATA veri bölgesidir. Bu alanda gerçek örnekleme verileri tutulur. DATA veri bölgesi alanları Tablo 4.3’te gösterilmiştir [11].

Tablo 4.3. Data veri bölgesi [11].

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

#### 4.4.2. Seslerin Ön İşlemden Geçirilmesi

Ses parçaları birbirlerine fazları, frekansları ve enerjileri uyumlu olacak şekilde birleştirilmelidir. Bu yüzden sesler veritabanına eklenmeden önce bu özelliklerinin birbirine uyumlu hale getirilmesi gerekmektedir. Bu işlem genellikle belli bir pencere fonksiyonu kullanılarak yapılır.

Sinyaller işlenmeden önce belli sayıda örnek içeren parçalara ayrılmaktadırlar. İşte bu parçaların her birine pencere adı verilmektedir. Pencere fonksiyonları, sinyal parçalarını çarpan fonksiyonlar olarak adlandırılmaktadır [38]. Bu fonksiyonlar yardımıyla sinyal parçalarının orta kısımlarının vurgulanması diğer kısımların yani pencerelerin başlangıcına

ve bitişine yakın kısımların söndürülmesi sağlanmaktadır. Bu sayede spektral sızıntı zayıflatılmış olur [38]. Eklemeli sentezleme kullanan sistemlerde genel olarak Hanning Pencere fonksiyonu kullanıldığı görülmüştür.

#### 4.4.2.1.Hanning Pencere Fonksiyonu

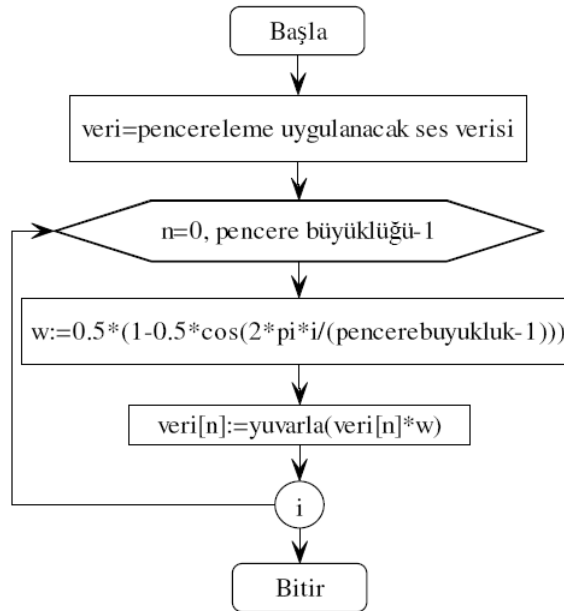
Hanning pencere fonksiyonunun formülü (Eşitlik 4.1' de) verildiği gibidir.

$$w[n]=0.5(1-\cos (2Pn/(N-1))) \quad (4.1)$$

Yukarıda verilen matematiksel formülün programlama dilindeki karşılığı şu şekilde olmaktadır:

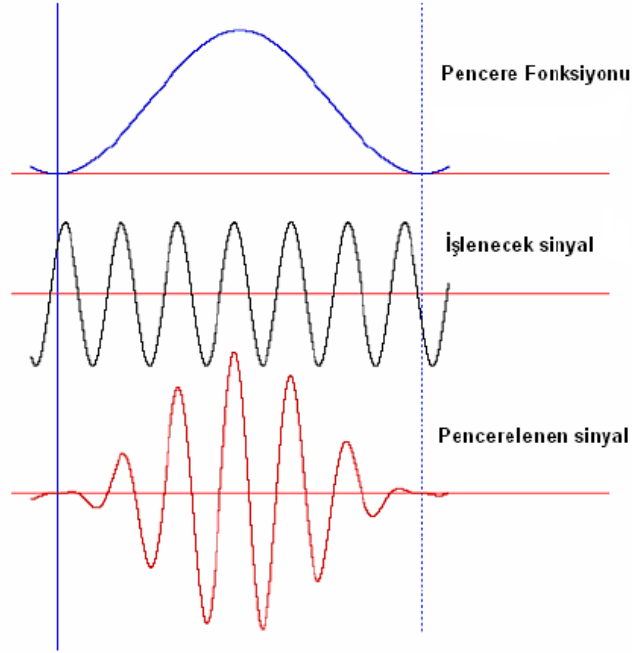
$$w=0,5*(1-0,5*\cos(2*pi*i/(pencerebuyukluk-1)))$$

Bu satır ile Hanning Pencere fonksiyonu hesaplanabilir ve bu fonksiyon tüm kaynak ses verisine uygulanabilecektir. Hanning Pencere işlemini gerçekleştiren alt programın akış seması Şekil 4.2'de gösterilmektedir.



Şekil 4.2. Hanning Pencere fonksiyonu akış seması

Hanning penceresinden geçirilmiş ses verileri konuşma sentezleme çalışması esnasında birleştirilen seslerin fazlarının ve frekanslarının eşit olması sağlayabilmektedir. Şekil 4.3'te örnek bir sinyalin pencere fonksiyonundan geçtikten sonraki görüntüsü verilmiştir.



Şekil 4.3. Örnek sinyalin Hanning Pencere fonksiyonu uygulandıktan sonraki görüntüsü

#### 4.5. Seslerin Veritabanına Kaydedilmesi

Türkçe alfabesi incelendiğinde üretilmesi gereken sekiz farklı hece türü bulunmaktadır. Bunlar en az bir, en fazla dört harften oluşur. Aslında öz Türkçe’de altı farklı hece tipi bulunmaktadır, ancak günümüzde diğer dillerden gelen ve dilimizce benimsenmiş birçok sözcük bulunmaktadır ve bu hece tipleri dahil edilmediği takdirde çoğu metin seslendirmesinde sorun yaşanacaktır [5,39]. Hece tabanlı oluşturulacak bir MKS sisteminde bulunması gereken hece tipleri ve sayıları şu şekilde özetlenebilir:

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

- V Tipinde Heceler: Tek sesli harften oluşan hecelerdir ve toplamda 8 adet V tipinde hece bulunmaktadır. (Örnek: a, e, ı, i, o, ö, u, ü)
- CV Tipinde Heceler: Sessiz harf + sesli harf şeklinde oluşan hecelerdir (Örnek: al, an, et, üç, öl). Matematiksel olarak  $21 \times 8 = 168$  adet CV tipinde hece bulunmaktadır.

- VC Tipinde Heceler: Sesli harf + sessiz harf şeklinde oluşan hecelerdir (Örnek: ba, ce, zi, ğü). CV hece tipinde olduğu gibi  $8 \times 21 = 168$  adet ses dosyası gerekmektedir.
- VCC Tipinde Heceler: Sesli harf + sessiz harf + sessiz harf şeklinde oluşan hecelerdir (Örnek: ilk, ürk, ast). Matematiksel olarak  $8 \times 21 \times 21 = 3528$  adet ses dosyası oluşması gerekmektedir. Ancak Türkçe diline ait kurallar göz önüne alındığında bu sayı indirgenebilir. Örneğin; “Aynı hecede iki ünsüz harf varsa bu ünsüz harf çifti "lç, lk, lp,lt, nç, nk, nt, rç, rk, rp, rs, rt, st, şt" olmalıdır [4, 5]. Kuralı ve yabancı kökenli sözcüklerde de bulunabilen “rz” sessizleri (örnek: tarz, ırz) ile daha çok ünlem içeren sesleniş kalıplarında görülen “yt” sessizleri (örnek: “hey”) iki ünsüz harfin bir arada bulunduğu durumlardır [5].
- CVC Tipinde Heceler: Sessiz harf + sesli harf + sessiz harf şeklinde oluşan hecelerdir (örnek: kal, tek, bit).  $21 \times 8 \times 21 = 3528$  adet ses dosyası gerekmektedir.
- CVCC Tipinde Heceler: Sessiz harf + sesli harf + sessiz harf + sessiz harf şeklinde oluşan hecelerdir (Örnek: Türk, sark, dört). Matematiksel olarak  $21 \times 8 \times 21 \times 21 = 74088$  adet ses dosyası gerekmektedir.
- CCV Tipinde Heceler: Sessiz harf + sessiz harf + sesli harf şeklinde oluşan hecelerdir (örnek: tra, spo, gri). Matematiksel olarak  $21 \times 21 \times 8 = 3528$  adet ses dosyası gerekmektedir. Yabancı kökenli sözcüklerde bulunabilen CCV tipi hece yapısı, sadece sözcük başlarında bulunabilir [5]. Yabancı kökenli sözcüklerin başında bulunabilen bu hece tipi sadece (br, bl, dr, fr, gl, gr, hr, kl, kr, pl, ps, tr) listesindeki iki sessizlerle başlayabilirler [5]. Bu durum göz önünde bulundurulduğunda oluşturulması gerek hece sayısı daha da düşecektir.
- CCVC Tipinde Heceler: Sessiz harf + sessiz harf + sesli harf + sessiz harf şeklinde oluşan hecelerdir (örnek: drop, klor, kros). Matematiksel olarak  $21 \times 8 \times 21 \times 21 = 74088$  adet ses dosyası gerekmektedir. CCV tipindeki heceler gibi sadece kelimenin başında bulunabilen bu heceler sadece (br, bl, dr, fr, gl, gr, hr, kl, kr, pl, ps, tr) sesleriyle başlayabilirler [5].

Toplam 8 hece türü için matematiksel olarak oluşturulacak hece sayısı

$$8 + 168 + 168 + 3528 + 3528 + 3528 + 74088 + 74088 = 159104$$

Türkçedeki imla kuralları göz önünde bulundurulduğunda ve hecelerin oluşmasını sağlayan kurallar dikkate alınarak bu sayıyı 7822 adede düşürmek mümkündür [5].

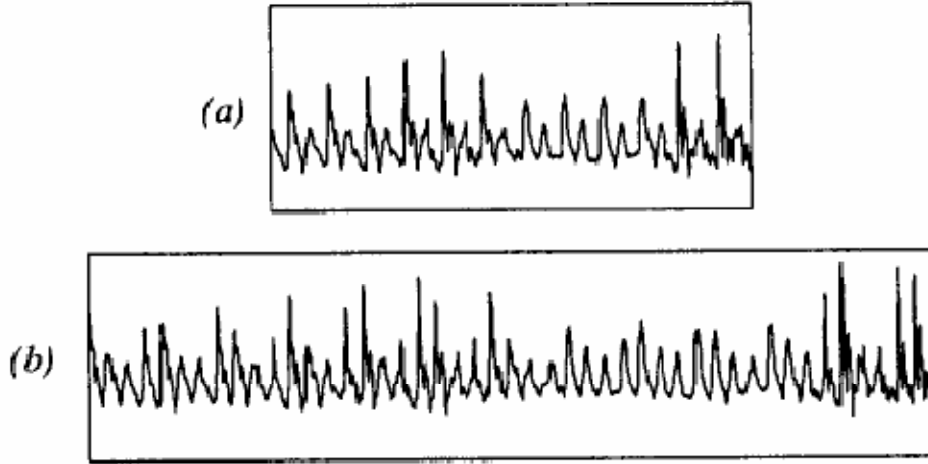


## 4.6. Seslerin Birleştirilmesi

Eklemeli metinden konuşma sentezleme çalışmalarında daha önceden kaydedilmiş farklı tonlamada ki sesleri birbirlerine istenilen ton ve süre ile bağlamak ve bir süreklilik kazandırmak için zaman ölçeği modifikasyonları kullanılmaktadır. Genel olarak bilinen 4 tür modifikasyon teknikleri bulunmaktadır. Bunlar; OLA, SOLA, TD-PSOLA ve WSOLA'dır.

### 4.6.1. Örtüşme-Ekleme Algoritması

Bu teknikte, Şekil 4.4'te de görüldüğü gibi zaman sıkıştırma ve genişleme işlemleri, örneklenmiş gerçek ses sinyalinin, pencerelenmiş granüllerin çıkarılması ve belli bir süre içinde yeniden düzenlenmesi ile gerçekleştirilir [8, 42].



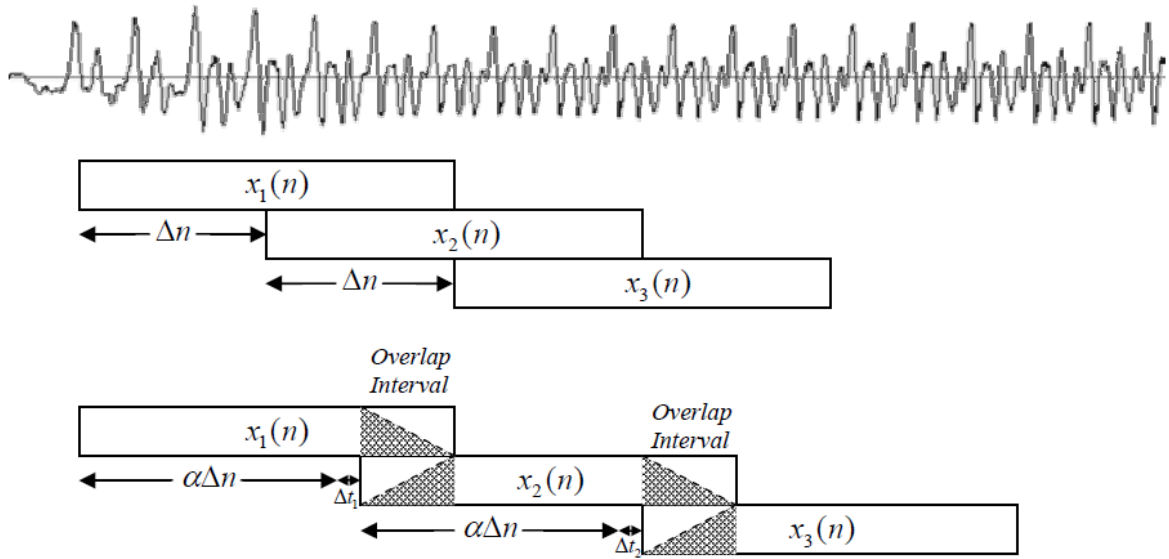
Şekil 4.4. OLA metodu ile zaman ölçekleme [42].

OLA, birbirlerine bağlı olan girdi sinyali parçacıklarının orijinal faz ilişkilerini yok eder ve ardından hizalanmamış sinyal parçacıkları arasında ara değerlendirme yaparak yeni çıktı sinyalini oluşturur. Bu işlem perde periyotlarında sinyal kalitesini etkileyen düzensizlik ve bozulmalara yol açar [42].

### 4.6.2. Eşzamanlı Örtüşme-Ekleme Algoritması

İlk olarak Roucos ve Wilgus tarafından tanımlanan eş zamanlı OLA metodu (SOLA) korelasyon tekniklerini temel alan zaman ölçeği sıkıştırma ve genişlemeye dayalı bir

algoritmadır [3]. Başlangıçta basit hesaplar gerektirdiğinden ve gerçek zaman uygulamalarına uygunluğundan dolayı bilgisayar tabanlı sistemlerde popüler olmuştur. Şekil 4.5'te görüldüğü üzere bu metotta girdi sinyali sabit uzunluklu çakışan bloklara bölünür ve her bir blok zaman ölçek faktörü  $\alpha$  sabitine göre kaydırılır. Daha sonra en yüksek karşılıklı öz ilinti değerine sahip ayrık zaman gecikmesi " $n\Delta t$ " örtüşme aralığı üzerinde araştırılır. Maksimum benzerliğin bulunduğu noktada, örtüşme bloklarına fade-in (derece derece açılma) ve fade-out (derece derece kararırma) fonksiyonları yardımıyla ağırlık değerleri verilir [8, 42].

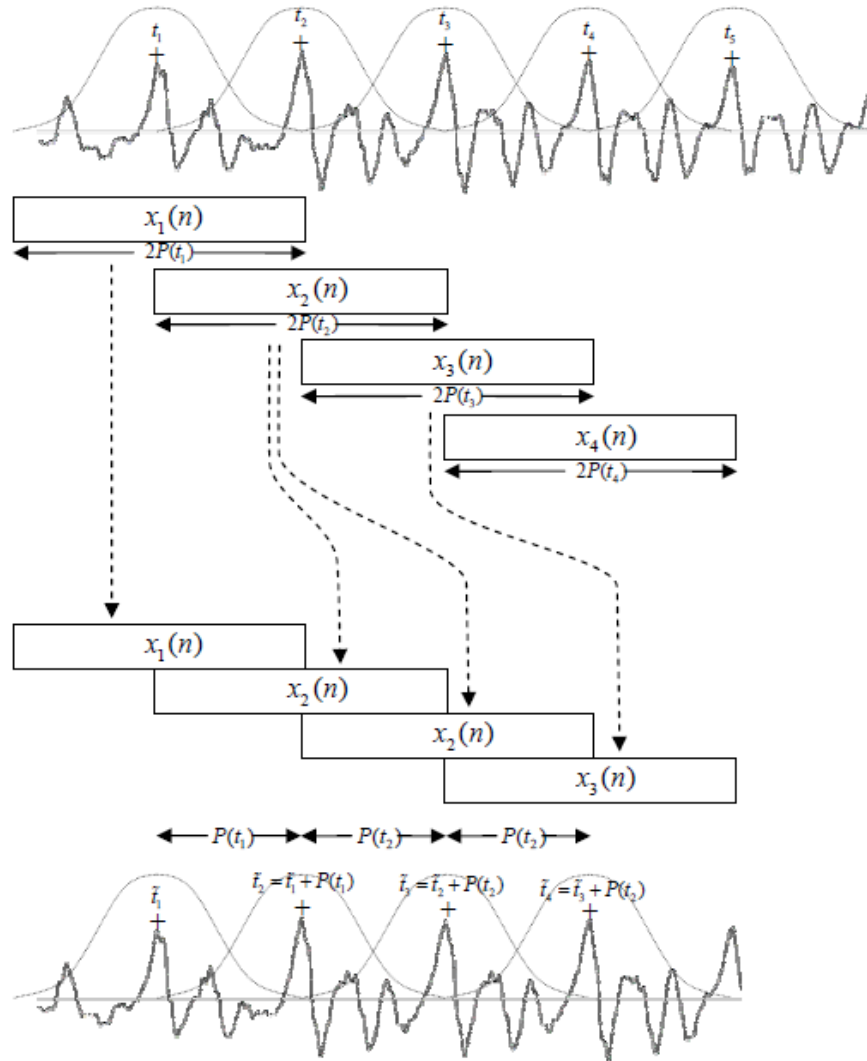


Şekil 4.5.SOLA metodu ile zaman ölçekleme [42].

Bu teknik sinyalin perde, büyüklük ve fazını korumasını sağlar. Makhoul, fade-in ve fade-out fonksiyonları olarak lineer ve tabanlı kosinüs fonksiyonlarını kullanmış ve lineer fonksiyon kullanmanın yeterli olduğunu göstermiştir [7]. Gürültülü ortamlarda sinyal kaydı yapılmış ise SOLA metodu OLA'ya göre daha iyi sonuç veren bir algoritmadır.

### 4.6.3. Zaman-Alan Senkronize Ton Örtüşme ve Ekleme Algoritması

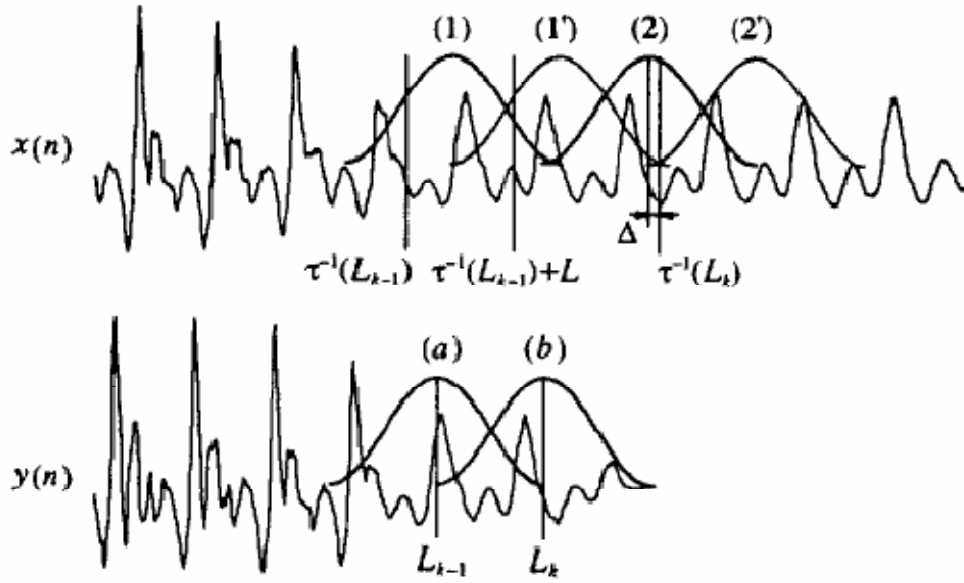
Sinyalin bir ton (örneğin, insan sesi ve monofonik müzik enstrümanları) aracılığıyla karakterize edildiği hipotezine dayanan TD-PSOLA SOLA algoritmasının bir varyasyonu olarak üretilmiştir [42]. Örtüştürme fonksiyonuna senkronize edilmiş bir ton uygulanarak yapılan bu teknikle yüksek kalitede zaman-alan modifikasyonu elde edilebilir. Bu yöntemin ana problemlerinden biri temel frekansın kayıp olduğu durumlarda sinyalin temel ton periyodunu tahmin etmede yaşanan sıkıntıdır. Algoritma iki bölüme ayrılabilir; Birincisi bölümde girdi sinyali analiz edilerek bölümlenmektedir. İkinci bölümde ise bölümlenen sinyali örtüştürme yöntemleriyle birleştirip ton eklemesi yapılmasıdır [42].



Şekil 4.6. TD-PSOLA metodu ile zaman ölçekleme ( $\alpha > 1$ ) [42].

#### 4.6.4. Dalga Şekli- Benzerlik Tabanlı Senkronize Örtüştürme ve Ekleme Algoritması

“Verhelst ve Roelands” tarafından geliştirilen teknik geleneksel SOLA tekniğinin bir çeşididir. WSOLA ile çapraz korelasyonla bağlanan bölümlerde orijinal sinyalde doğal sürekliliğe maksimum benzerlik sağlamaya çalışılmaktadır (Şekil 4.7) [42]. Bu işleme tekniğinde kullanılan temel prensip, konuşma sinyali içerisinde belirli zaman periyotları boyunca dalga şekli benzerliklerinin ortaya çıkarılarak ses örneklerinin azaltılmasıdır [42].



Şekil 4.7. WSOLA metodu ile zaman ölçekleme [42].

Sonuç olarak PSOLA yönteminde perde işaretleri referans alınır ve süre değiştirme, perde frekansı modifikasyonu gibi işlemler perde eş zamanlı olarak yapılır. Böylece ses parçaları birbirine faz uyumlu bir şekilde eklenmiş olur [3, 42, 43].

WSOLA yönteminde ise; dalga şekli benzerliği (maksimum çapraz ilinti) referans alınır. WSOLA'da perde işaretlerine gerek yoktur. Bu yönden PSOLA'ya göre avantajlıdır. PSOLA yöntemiyle süre ve perde frekansı ayarlamaları; WSOLA yöntemiyle ise yalnızca süre ayarlaması yapılabilir [3].

## 5. METİNDEN KONUŞMA SENTEZLEME UYGULAMALARI

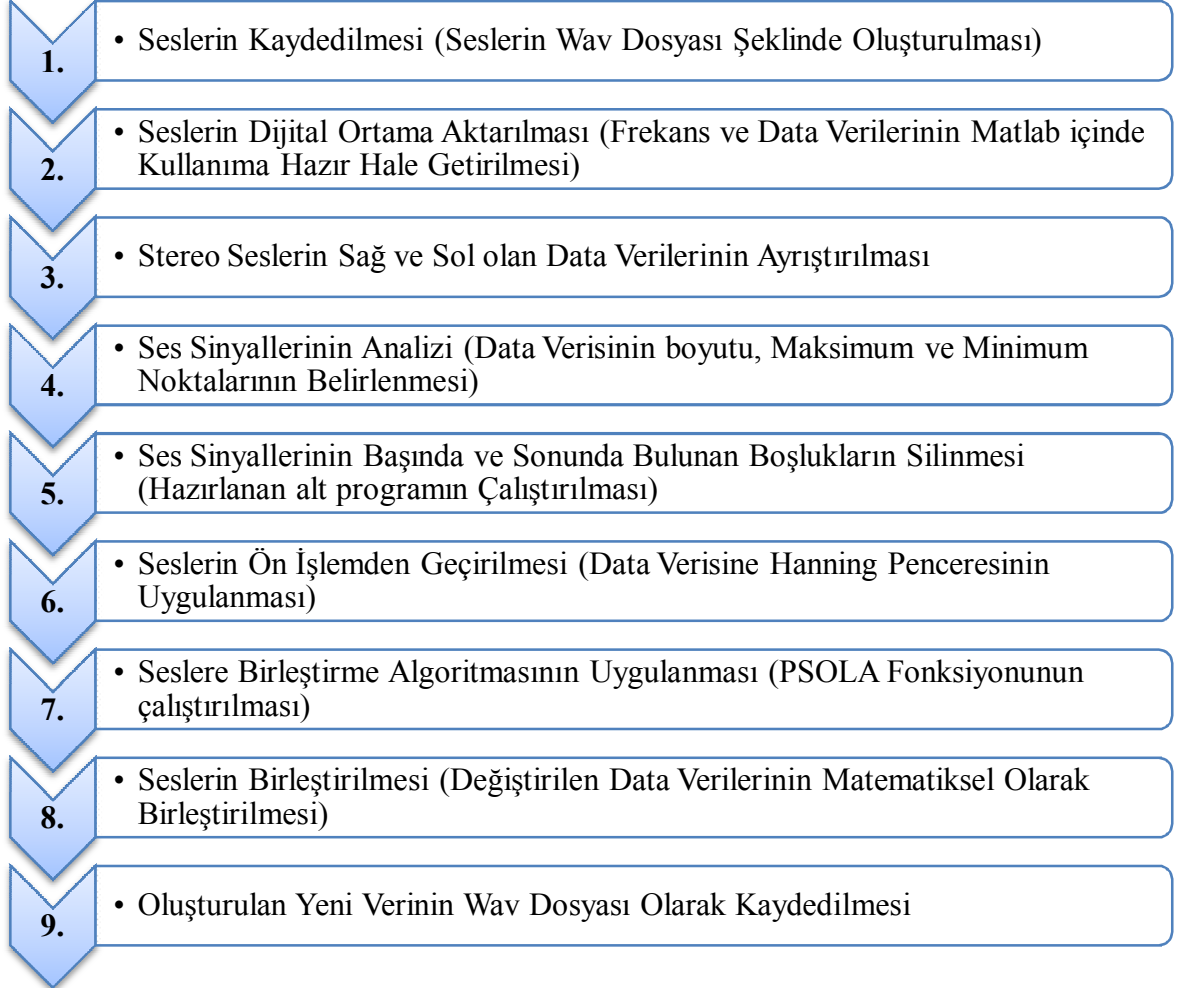
Bu tez çalışmasında 2 adet Metinden Konuşma Sentezleme sistemi tasarlanmıştır. Bu sistemler “Matlab” ve “C#” programlama dilleri kullanılarak hazırlanmıştır.

Matlab’da hazırlanan sistem de, Türkçe dilinin yapısından dolayı Eklemeli Sentezleme yöntemi kullanılmıştır. Türkçe kelimelerin hecelerden oluştuğu ve tüm kelimelere yeni heceler bağlanarak yeni kelimelerin oluşturulduğu göz önünde bulundurulup eklenecek parça olarak heceler seçilmiştir. Bu sistem içinde heceler ayrı ayrı ses dosyası halinde kaydedilip ve sonrasında bu ses dosyalarının sayısal ortama geçirilerek filtreleme, çıkarma, ekleme ve birleştirme gibi bir takım işlemlerden geçirilmesi sağlanmıştır. Ses dosyaları üzerindeki tüm bu işlemler Matlab yardımıyla istenilen şekilde yapılabilmektedir. Bu yüzden sistem alt yapısı olarak Matlab programlama dili kullanılmıştır. Yapılan bu uygulama örnek bir uygulama olup 3 adet hecenin “PSOLA” algoritması kullanılarak birleştirilmesiyle oluşturulmuştur. Matlab sinyal işleme uygulamalarında oldukça başarılı olmasına rağmen çoklu ortam ve nesne tabanlı uygulamalarda diğer programlama dillerine göre kullanıcıyla etkileşimi daha azdır. Bu sebeple çoklu ortam uygulamalarında daha az kullanılmaktadır.

Diğer hazırlanan sistem ise “.Net Framework 4.0” tabanlı olarak “C#” dili yardımıyla hazırlanmıştır. Şuan en popüler programlama dillerinden biri olan C# .net’in içerisinde bulunan tüm uygulamalara sorunsuz erişebilmektedir. Bu yüzden hazırlanan sistem C# programlama dili kullanılarak oluşturulmuştur. Bu sistem de .Net öğelerinden “Microsoft System Speech Application ” kullanılarak konuşma sentezleme işlemi gerçekleştirilmiştir. Fonemleri seslendirerek sentezleme yapabilen sistemde ses veritabanı olarak “Mbrola” projesiyle hazırlanan Türkçe-Erkek ve Türkçe-Kadın fonem ses veritabanları kullanılmıştır.

## 5.1. Hece Tabanlı Metinden Konuşma Sentezleme Sistemi

Hece Tabanlı MKS sisteminde sırasıyla yapılan işlemler Şekil 5.1’de blok şema halinde gösterilmiştir. Matlab tabanlı hazırlanan sistemde bu işlem basamakları sırasıyla izlenmiştir.



Şekil 5.1. Hece tabanlı TMKS sisteminin blok şeması

### 5.1.1. Seslerin Kaydedilmesi

Bir metni seslendirebilmek için o metni oluşturan metin parçalarının sesli karşılığının bilgisayar ortamında bulunması gerekmektedir. Bu çalışmada seslendirilecek metin parçası olarak hece birimi seçilmiştir. Hecelerin seslendirilerek bilgisayar ortamına aktarılması için işletim sistemi üzerinde bulunan ses kaydedicisi dışında ücretsiz olarak erişilebilen birçok

ses kaydetme programı internet ortamında bulunmaktadır. Bu çalışmada hecelerin seslendirilerek bilgisayar ortamına aktarılması için “<http://www.freesoundrecorder.net/>” sitesinden ücretsiz olarak indirilebilen “Free Sound Recorder 9.4.1” yardımcı programı kullanılmıştır.



Şekil 5.2. FreeSound Recorder programı ekran görüntüsü

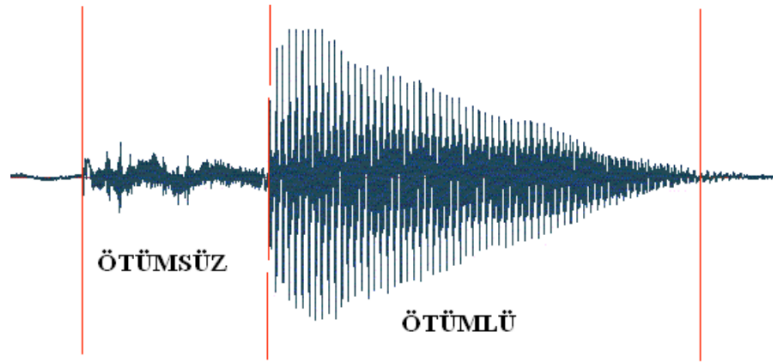
Heceleri kaydetmek için ses dosya türleri içinde en sade ve en işlenebilir format olan wav formatı kullanılmıştır. Bu format için 44100 kHz örnekleme frekansı ve 16 bit stereo çift yönlü olarak ses kalitesi seçilmiştir.

### 5.1.2. Seslerin Data Bölümünün Elde Edilmesi

“Wav” dosyaları Matlab programlama dili içerisinde kolayca dahil edilebilmektedir. Bu işlem için “wavread” komutu kullanılabilir. Bu komutla bir “wav” dosyası matlab da açıldığında 3 tip parametre sisteme değişken olarak alınabilir. Bunlar veri kısmı, örnekleme frekansı ve her örnek için bit sayısı değerleridir. Veri kısmında sesi oluşturan bilgiler bulunmaktadır yani sesin oluşmasını sağlayan matematiksel değerler bulunmaktadır. Sesin mono veya stereo olmasına bağlı olarak tek boyutlu veya çift boyutlu bir matris şeklindedir. Örnekleme frekansı o sese ait frekansı vermektedir. Yani sesin saniyede kaç kez örneklenecek ses verisinin oluşturulduğunu gösteren kısımdır. Bit sayısı

ise sesin kaydedilme yöntemine bağlı olarak 8 veya 16 değerlerini alabilmektedir. Hazırlanan uygulamada veri kısmı ve frekans kısmı aktif olarak kullanılacaktır. Bit sayısının bu uygulamada bir işlevi yoktur.

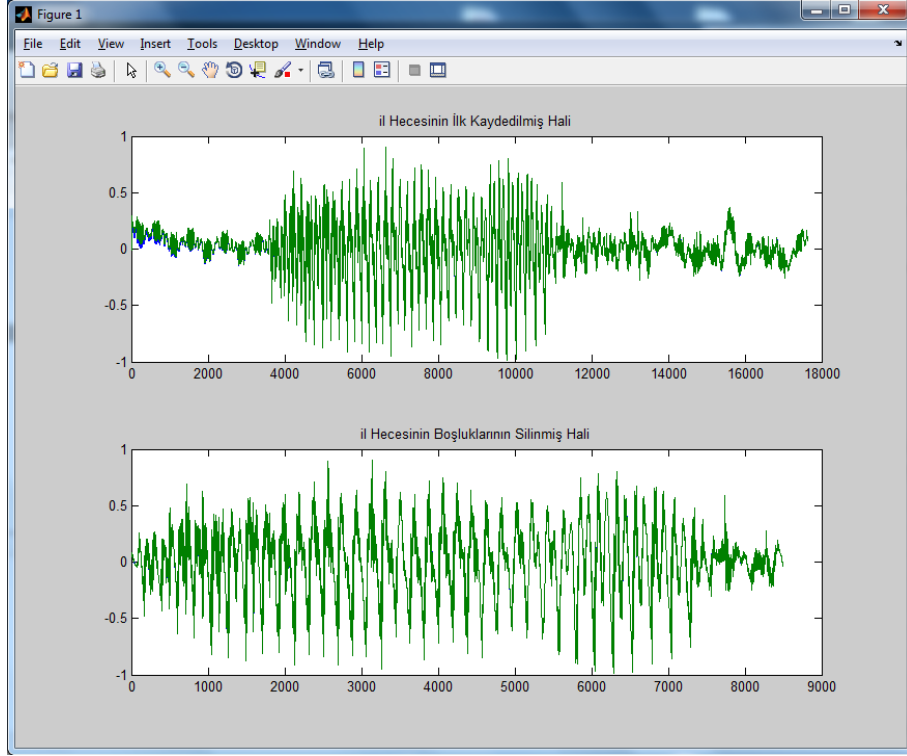
İnsan konuşmasında temelde iki farklı ses vardır. Bunlar ötümlü ve ötümsüz olarak adlandırılmaktadır [11]. Ötümlü ses zaman çerçevesinde incelenirse periyodik karakteristik gösterdiği görülmektedir. Diğer taraftan ötümsüz ses periyodik olmayan bir davranış sergilemektedir. Şekil 5.3'te ötümlü ve ötümsüz seslerin dalga şekli görülmektedir [11].



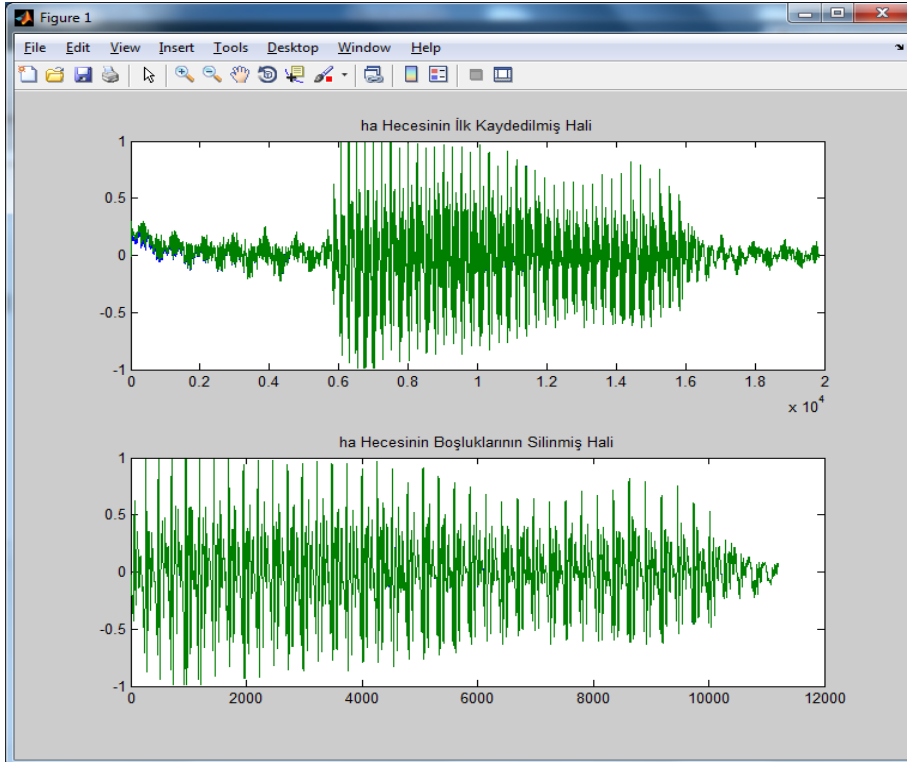
Şekil 5.3. Ötümsüz ve Ötümlü Ses [11]

Sesler kaydedilirken seslerin başında ve sonunda ötümsüz (ses olmayan) kısımlar ile beraber kaydedildiği görülmüştür. Bu sorunun giderilmesi için çok sayıda deneme yapılmasına rağmen mutlaka seslerin başında ve sonunda bu boşluklar oluşmaktadır. Bu sorunun giderilmesi için seslerin başındaki ve sonundaki boşlukların ses verisinden çıkarılması gerekmektedir. Bu işlem için Matlab programlama dili içerisinde hazırlanan bir alt program kullanılmıştır. Bu alt programda data verisi içinde bir eşik değeri seçilmiştir. Bu değer sesin oluşabilmesi için data kısmındaki minimum değer olarak düşünülebilir. Sonrasında data verisi bu eşik değere bağlı olarak bir karşılaştırmadan geçer. Bu karşılaştırma esnasında datanın başında veya sonunda bulunan ve bu eşik değerinin altında kalan veriler tüm veriden çıkarılarak verinin sadece sesteki oluşması sağlanmıştır. Oluşturulan yeni veri aynı örnekleme frekansı kullanılarak yeni ses dosyası olarak kaydedilmektedir. Şekil 5.4, 5.5 ve 5.6'da örnek ses verilerinin bu işlem öncesi ve sonrasındaki grafiksel görüntüsü verilmiştir.

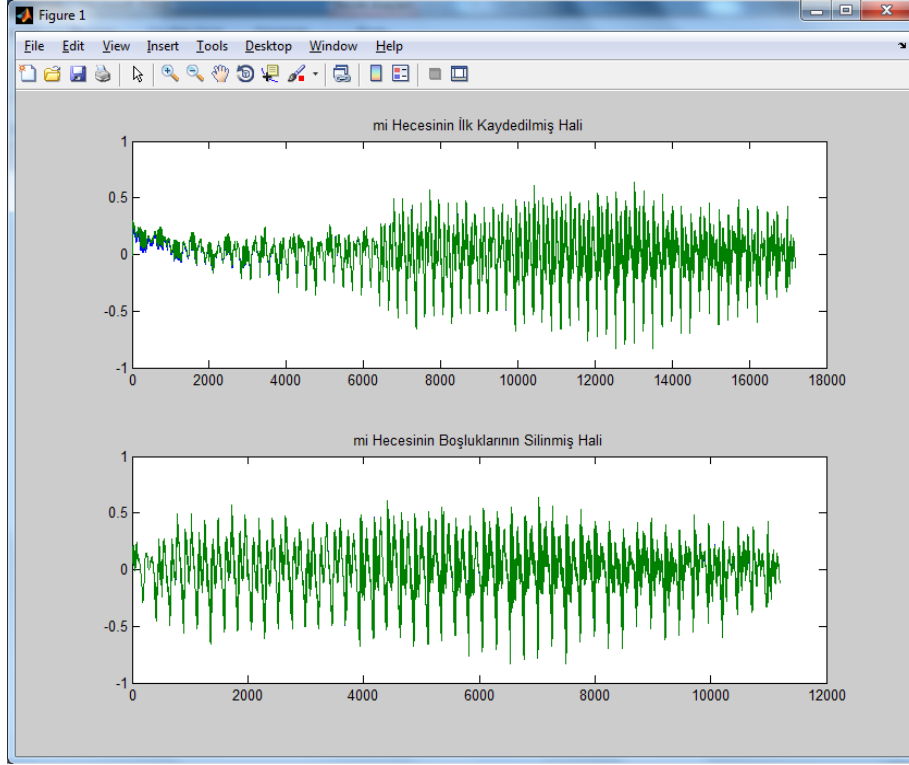




řekil 5.4. "il" Sesinin data verisinin elde edilmesi



řekil 5.5. "ha" sesinin data verisinin elde edilmesi



Şekil 5.6. “mi”sesinin data verisinin elde edilmesi

### 5.1.3. Seslerin Ön İşlemden Geçirilmesi

Birleştirilecek seslerin diğer seslerle uygun bir biçimde birleştirilebilmesi için frekanslarının ve genliklerinin aynı olması gerekmektedir. Bunun için de birtakım ön işlemlerden geçmesi gerekmektedir. Bu işlem için farklı yöntemlerin olduğu literatür araştırmasında görülmüştür. Bu çalışmada ise sesler Hanning penceresinden geçirilerek seslerin birleştirilmeden önceki son haline gelmesi sağlanmıştır. Bu işlem için Matlabda oluşturulan kodlar Şekil 5.7’de verilmiştir.

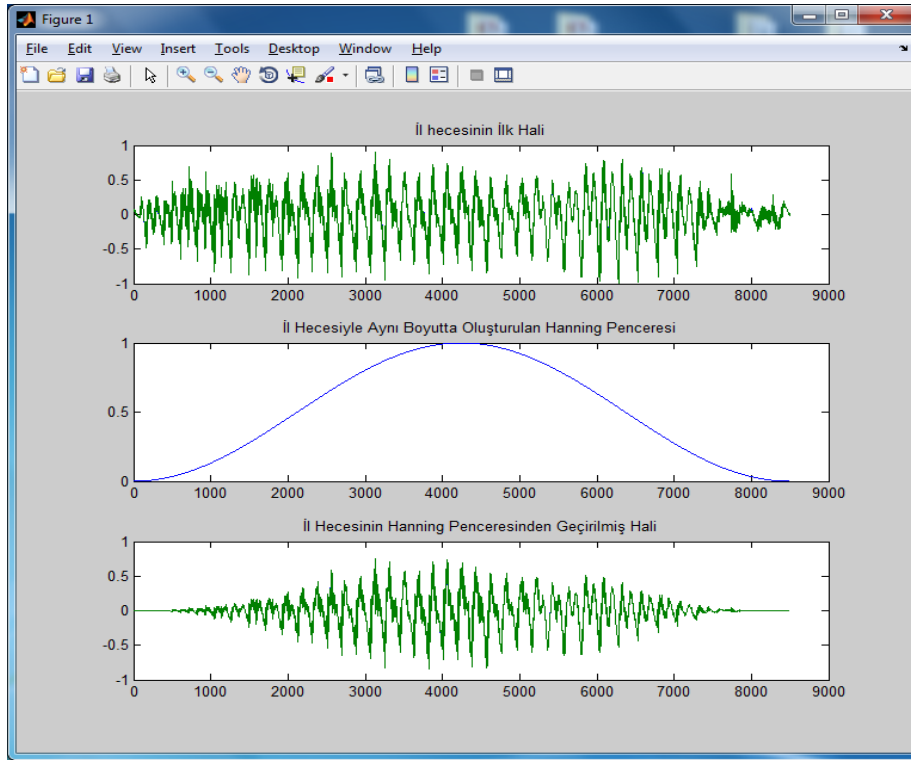
```

%% HANNING PENCERELEME YÖNTEMİ
[ses, fs] = wavread('il.wav');           // Wav dosyasının programa aktarılması
boyut=length(ses);                       // Ses verisinin boyutunun belirlenmesi
pencere=hann(boyut);                     // Aynı boyutta hanning penceresinin oluşturulması
ses_sag=ses(:,1);                         // Stereo sesin sağ bölümünün ayrıştırılması
ses_sol=ses(:,2);                         // Stereo sesin sol bölümünün ayrıştırılması
ses_sag=ses_sag.*pencere;                 // Pencere fonksiyonunun sağ bölüme uygulanması
ses_sol=ses_sol.*pencere;                 // Pencere fonksiyonunun sol bölüme uygulanması
yenises=[ses_sag ,ses_sol];               // Sağ ve sol bölümlerin birleştirilmesi
wavwrite('sesyeni', yenises, fs)         // Ses verisinin wav olarak kaydedilmesi

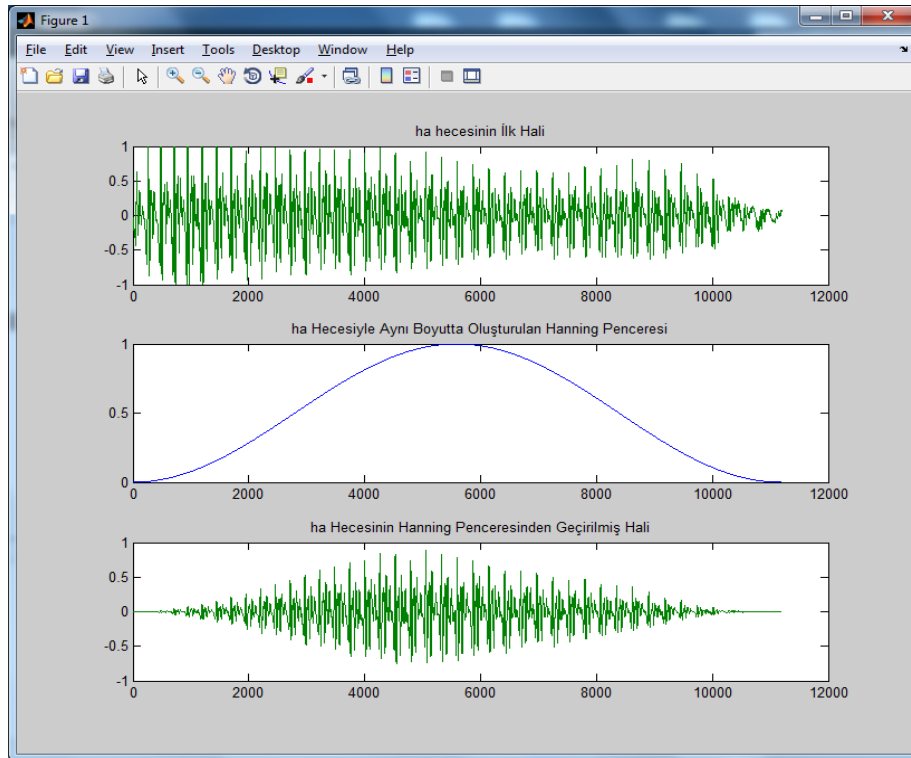
```

Şekil 5.7. Hanning pencereleme yöntemi örnek kodları

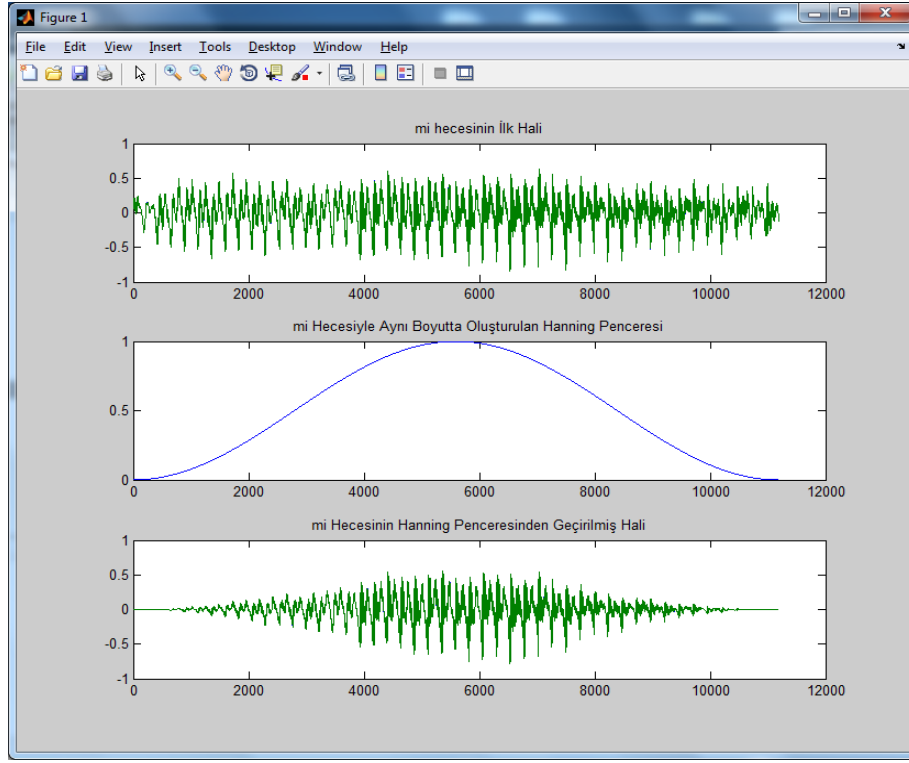
Aşağıda gösterilen şekillerde ise örnek hecelerin kaydedildiği ilk hali ve Hanning penceresinden geçtikten sonraki hali grafiksel olarak verilmiştir.



Şekil 5.8. “il” Sesinin Hanning penceresinden geçirilmesi



Şekil 5.9. “ha” Sesinin Hanning penceresinden geçirilmesi



Şekil 5.10. “mi”Sesinin Hanning penceresinden geçirilmesi

Hanning penceresinden geçirilerek bir filtre uygulanan sesler birleştirilmek için son haline getirilmiştir. Bu haliyle tekrar wav dosyası olarak kaydedilen sesler, MKS sisteminde kullanıma hazır hale gelmiştir.

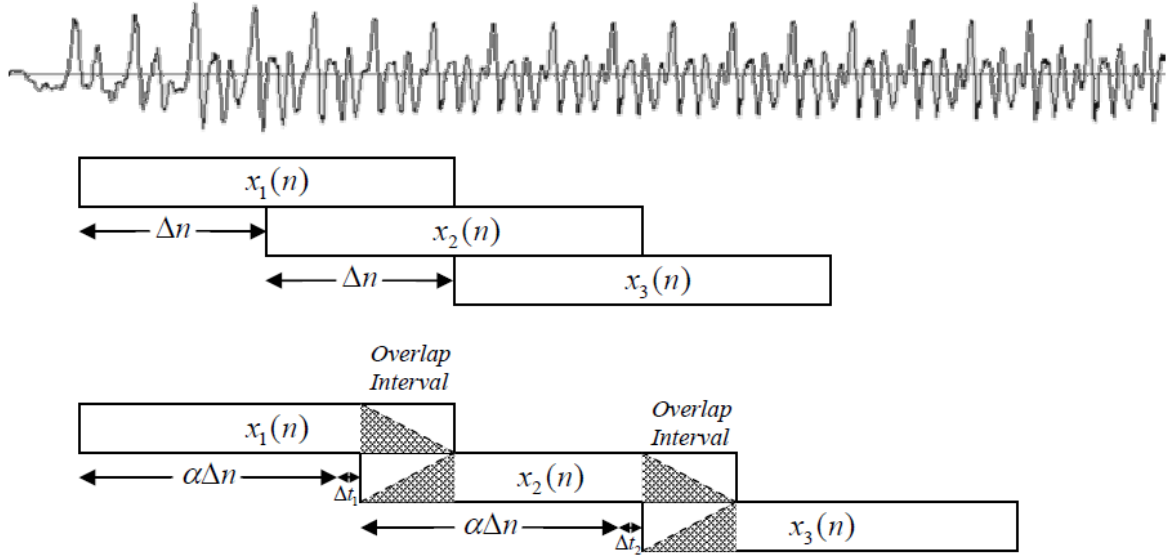
#### 5.1.4. Seslerin Birleştirilmesi

Ön işlemden geçirilerek son haline gelen heceler uygulama içerisinde birleştirilerek kelimeleri seslendirmeyi sağlar. Sesleri birleştirmek için kullanılan bir kaç yöntem vardır. Bu yöntemler Örtüşme-Ekleme Algoritması (Overlap-Add, OLA) ve bunun türevleri olan;

- SOLA (Eşzamanlı Örtüşme-Ekleme Algoritması)
- TD-PSOLA (Zaman-Alan Senkronize Ton Örtüşme ve Ekleme Algoritması)
- WSOLA (Dalga Şekilli- Benzerlik Tabanlı Senkronize Örtüştürme ve Ekleme Algoritması)

Bu yöntemler detaylı bir şekilde dördüncü bölümde anlatılmıştır. Türkçe metin seslendirme üzerine yapılan çalışmalar araştırıldığında en çok kullanılan yöntemlerin

SOLA ve TD-PSOLA algoritmaları olduğu görülmüştür. Hazırlanan bu sistemde ise sesleri birleştirmek için SOLA algoritması kullanılmıştır. Şekil 5.10'da SOLA algoritmasının çalışma mantığı grafiksel olarak gösterilmiştir [42].



Şekil 5.11. SOLA metodu ile zaman ölçekleme [42].

Bu yöntemde birleştirilecek seslerin belirli kısımlarına bir filtre uygulanarak ses geçişlerinin daha yumuşak olması sağlanabilmektedir. Filtre uygulanan kısımlar üst üste örtüştürülerek bu yöntem uygulanmaktadır. “Fade-in” ve “Fade-out” efekti uygulanacak bölümleri sesin %30’ u olacak şekilde belirlenmiştir. Hazırlanan sistemde kullanılan sesler %30 ilk bölüm, %40 esas bölüm ve %30 son bölüm olarak parçalara bölünmüştür. Bu işlem için Şekil 5.12'deki kodlar kullanılmıştır.

```
[ses,fs] = wavread('il.wav'); //sesin dijital ortama aktarılması
sesl =ses(:,1); //ses verisinin sol bölümünün ayrılması
sesr =ses(:,2); //ses verisinin sağ bölümünün ayrılması
boyut=length(ses); //verinin boyutunun belirlenmesi
x=ceil(boyut*30/100); //"Fade in" ve "Fade out" boyutunun belirlenmesi
anaboyut=ceil(boyut*40/100); //aynı kalacak verinin boyutunun belirlenmesi
```

Şekil 5.12. Sesin bölüm boyutlarının belirlenmesi için kullanılan kodlar

```

%% sesi bölümlerine ayırma
% sesin ilk bölümü
seslilk=sesl(1:x, 1);
sesrilk=sesr(1:x, 1);
-----
% sesin ana bölümü
seslana=sesl(x:anaboyut,1);
sesrana=sesr(x:anaboyut,1);
-----
%sesin son bölümü
t=x+anaboyut;
seslson=sesl(x:end, 1);
sesrson=sesr(x:end, 1);

```

Şekil 5.13. Sesin bölümlere ayrılması için kullanılan kodlar

İlk hecenin %30'luk son kısmına Fade-out (yavaş yavaş kaybolma) efekti verilmesi gerekmektedir. Bu işlem için o boyutta bir ters rampa fonksiyonu oluşturulmuştur. Oluşturulan bu fonksiyonla sesin son kısmı matris çarpımı yapılarak kendisinden sonra gelen sesle birleştirilmek üzere hazır hale gelmiştir. Bir sonraki sesin ilk bölümüne ise yani %30'luk bölüme fade in (yavaş yavaş belirme) efekti verilmiştir. Bunun için programlama ortamında da bu bölümün boyutunda bir rampa fonksiyonu oluşturulmuştur. Bu fonksiyonla sesin ilk kısmını matris çarpımı yaparak sesin başka bir hecenin sonuna eklenebilmesi için uygun hale gelmesi sağlanmıştır.

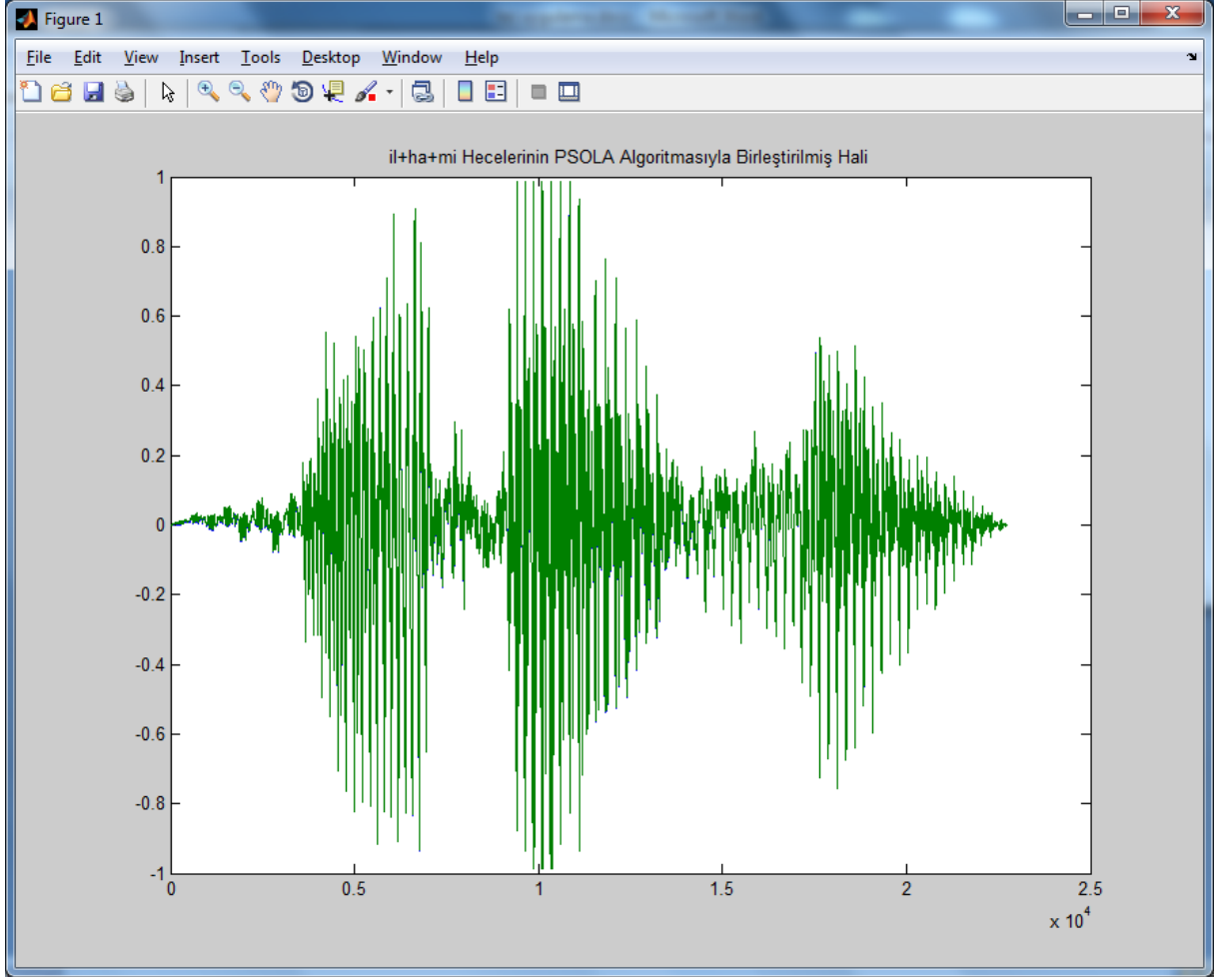
```

ramp=1/x:1/x:1; // rampa oluşturulması (Fade in efekti için)
ramp=transpose(ramp); // veri ile çarpılabilmesi için "transpose"
// işleminin yapılması
-----
tersramp=1/x:1/x:1; // rampa oluşturulması (Fade out efekti için)
tersramp=tersramp(x:-1:1); // Tersini alarak ters rampa oluşturulması
tersramp=transpose(tersramp); // veri ile çarpılabilmesi için "transpose"
// işleminin yapılması
-----
% sesin ilk bölümü fadein
seslilk=seslilk.*ramp; // Ses datasının, sol verisinin ilk kısmına
// fade in efekti verilmesi
sesrilk=sesrilk.*ramp; // Ses datasının, sağ verisinin ilk kısmına
// fade in efekti verilmesi
-----
% sesin son bölümü fadeout
seslson=seslson.*tersramp; // Ses datasının, sol verisinin son kısmına
// fade out efekti verilmesi
sesrson=sesrson.*tersramp; // Ses datasının, sağ verisinin son kısmına
// fade out efekti verilmesi

```

Şekil 5.14. Sesin bölümlerine efektlerin verilmesi için kullanılan kodlar

Şekil 5.15'de "ilhami" kelimesini oluşturan hecelerin PSOLA algoritması kullanılarak birleştirilmesi grafiksel olarak gösterilmektedir.



Şekil 5.15. “il+ha+mi” Hecelerinin PSOLA algoritmasıyla birleştirilmiş halinin grafiksel gösterimi

## 5.2. Difon Tabanlı Metinden Konuşma Sentezleme Sistemi

Bu uygulama da diğer uygulamadan farklı olarak daha çok son kullanıcıya hitap edecek bir MKS sistemi kullanılmıştır. Nesne tabanlı olarak hazırlanan uygulama kullanım açısından oldukça basit bir şekilde tasarlanmıştır. Ayrıca derlendikten sonra tüm Windows işletim sistemleri altında kolayca çalıştırılabilmektedir. Fonem seslendirmesi metoduyla çalışan sistemde Mbrola projesiyle hazırlanan ses veritabanları kullanılmıştır.

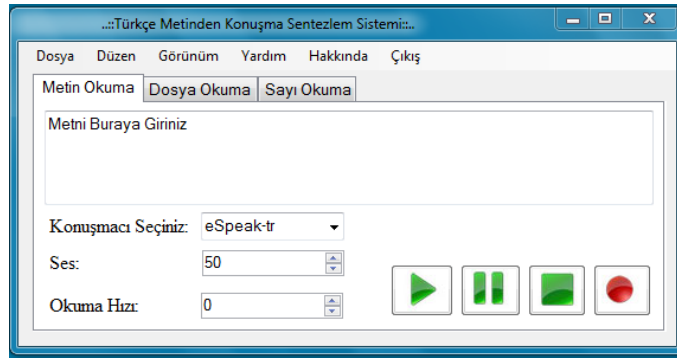
### 5.2.1. Çalışma Ortamı

Uygulama hazırlanırken .Net Framework 4.0 tabanlı olarak C# dili kullanılmıştır. Şuan en popüler programlama dillerinden biri olan C# .net'in içerisinde bulunan tüm uygulamalara sorunsuz erişebilmektedir. Bu uygulamada .Net öğelerinden Microsoft

System Speech Application kullanılarak konuşma sentezleme işlemi gerçekleştirilmiştir. Konuşma dili olarak Mbrola projesinde hazırlanan Türkçe-Erkek ve Türkçe-Kadın fonem ses veritabanları kullanılmıştır.

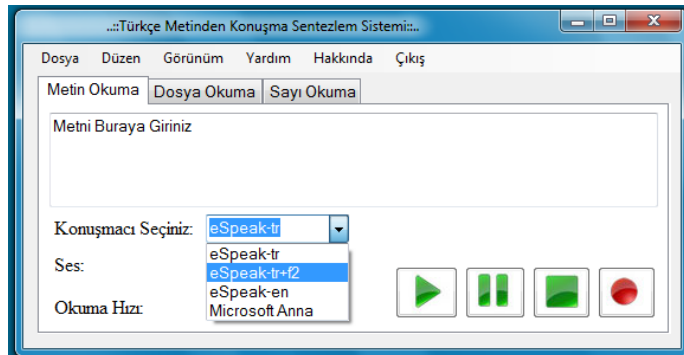
### 5.2.2. Uygulama İçeriği

Oluşturulan programın ekran görüntüsü Şekil 5.16’da görülmektedir. Uygulamanın karşılama ekranında metnin girilmesi için metin kutusunun yanında konuşma özelliklerinin ayarlanabileceği diğer nesnelere görülmektedir. Konuşmanın başlatılması, durdurulması, devam ettirilmesi için gerekli butonların yanında sentezlenen metni wav dosyası şeklinde kaydetmek için bir kayıt düğmesi de bulunmaktadır.



Şekil 5.16. Uygulamanın ekran görüntüsü

Konuşmacı seçimi için programa combobox eklenmiştir (Şekil 5.17). Sistemde yüklü olan dilleri görüntüleyip o an için hangi dil kullanılacaksa buradan seçilebilir. Uygulamada Windows’un kendi standart sesi olan “Microsoft Anna” dışında Mbrola projesinde hazırlanan İngilizce ile Türkçe erkek ve kadın sesi programa eklenmişti.

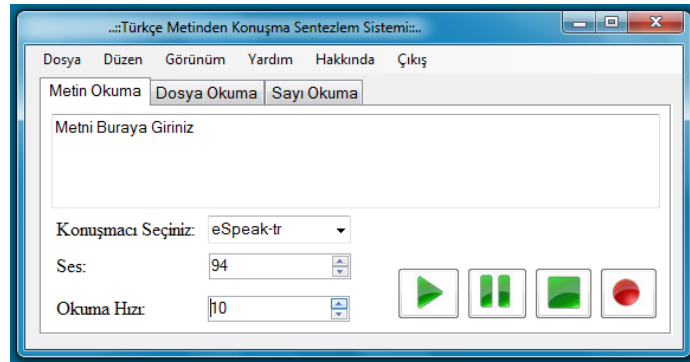


Şekil 5.17. Konuşmacı seçimi



Uygulamaya ses (volume) ayarı için “numericupdown” nesnesi eklenmiştir. Konuşma sesinin maksimum seviyesi 100 minimum seviyesi ise 20 olarak ayarlanmıştır. Ses değeri klavyeden girilebileceği gibi artırma ve azaltma tuşları sayesinde de ayarlanabilir.

Sentezlenen konuşmanın hızı (rate) için “numericupdown” tuşu kullanılmıştır (Şekil 5.18). Okuma hızı konuşmacının kelimeleri ne kadar hızlı veya ne kadar yavaş okumasını sağlamak için kullanılmıştır. Maksimum seviyesi 10 minimum seviyesi ise -10’dur. Okuma hızı 10 da iken en hızlı -10 da iken en yavaş şekilde konuşmayı sentezlemektedir.



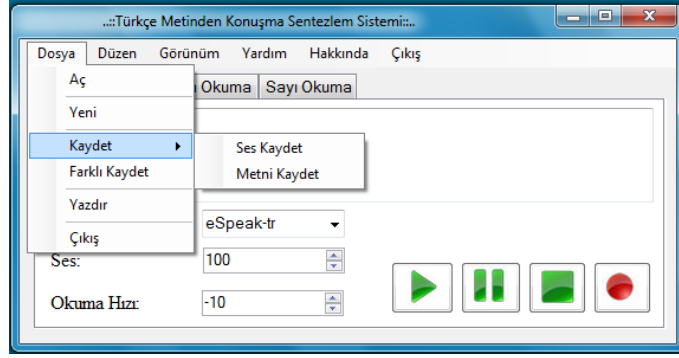
Şekil 5.18. Okuma hızının ayarlanması

Dosya menüsünde klasik Office yazılımlarında bulunan dosya öğeleri bulunmaktadır (Şekil 5.19). Aç öğesi bilgisayarda kayıtlı metin dosyasını açmak için kullanılmaktadır. Aç öğesi tıklandığında diyalog kutusu açılıp bir metin dosyası seçimini beklemektedir. Uygulama txt uzantılı dosyaları sorunsuz bir şekilde açmaktadır. Metin dosyası seçildikten sonra uygulama dosya okuma bölümüne geçmektedir.

Yeni öğesi seçildiği zaman programın tüm ayarlarını sıfırlayıp yeni bir uygulama açmaktadır. Uygulamanın default ayarları konuşmacı olarak Türkçe erkek sesi, ses ayarı olarak 90, okuma hızı ise 0 olarak belirlenmiştir.

Kaydet öğesi sesi ve metni kaydetmek için kullanılmaktadır. Yazılan metni kaydederken “txt” uzantılı olarak kaydetmektedir. Sesi kaydet öğesi seçildiğinde ise sentezlenen metnin sesini kaydetmektedir. Bu sesin uzantısı ise wav şeklinde olmaktadır.

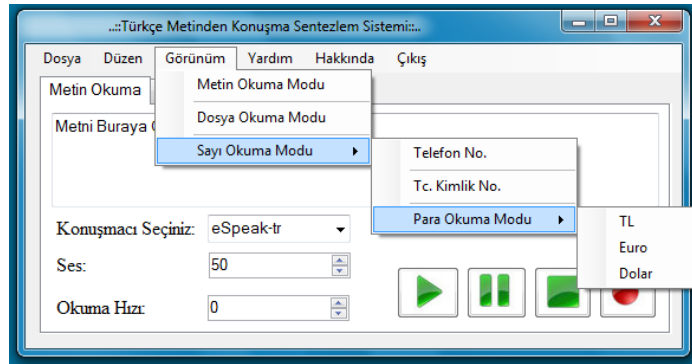
Yazdır öğesi yazılan metnin yazdırılması sırasında kullanılmaktadır.



Şekil 5.19. Dosya menüsü

Düzen menüsü içerisinde kes, kopyala, yapıştır öğelerinin dışında yazılan metnin seçilmesini ve konuşmacı seçeneğini değiştirmek için kullanılacak öğeler bulunmaktadır.

Görünüm menüsü içerisinde uygulamanın konuşma sentezleme uygulamaları seçilmektedir (Şekil 5.20). Yazılacak yazının sentezlenmesi için metin okuma modu seçilmelidir. Dosya okuma modunda ise metin dosyalarının okunacağı bölüm açılacaktır. Sayı okuma modunda telefon, T.C. kimlik no ve para birimi tipinde olan rakamların okunması için seçilecek öğeler bulunmaktadır. Para birimi okuma modu içerisinde TL, Euro ve Dolar para birimlerinin okunması için seçenekler bulunmaktadır.



Şekil 5.20. Görünüm menüsü

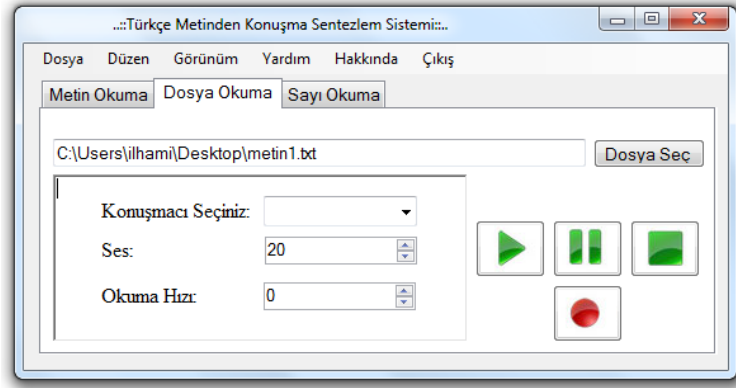
Yardım menüsünde program kullanımı öğesi bulunmaktadır. Programın kullanımı için gerekli bilgilere buradan ulaşılabilir. Program hakkında öğesi sayesinde programın lisans bilgileri ile birlikte iletişim bilgileri de yer almaktadır. Çıkış menüsü sayesinde programdan çıkış yapılabilir.

### 5.2.3. Metnin Seslendirilmesi

Program ilk açıldığında açılan mod metin okuma modu olmaktadır. Metin okuma modunda “textbox” içine yazılan metin seslendirilmektedir. Konuşmacı seçimi, ses ve okuma hızı da bu ekrandan ayarlanabilmektedir. “oynat, duraklat ve durdur” butonlarının yanında kayıt butonu da bu ekranda bulunmaktadır. oynat tuşuyla başlatılan sentezleme işlemi duraklat ile durdurulup kaldığı yerden tekrar başlatılabilir. Durdur butonu ise sentezlemeyi tamamen durdurabilir. Kayıt butonuna basıldığında sentezlenen konuşma ses dosyası olarak kaydedilmektedir. Bu işlem için diyalog kutusu açılıp wav dosyasını bilgisayarda nereye kaydedilmesini istendiği sorulacaktır.

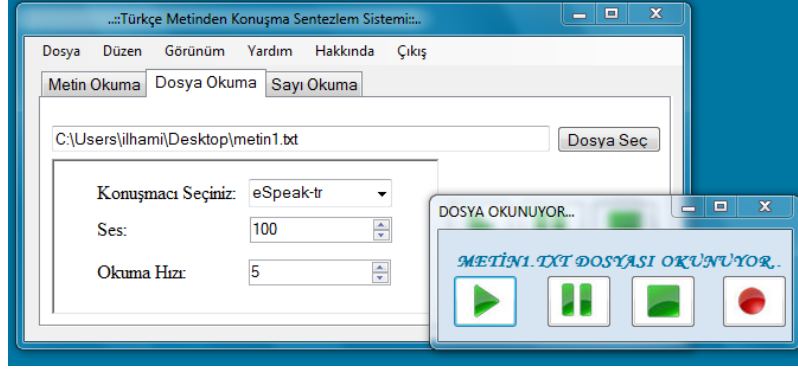
### 5.2.4. Dosyaların Seslendirilmesi

Dosya okuma modu Şekil 5.21’de görünmektedir. Metin okuma modunda bulunan öğelerin dışında bilgisayardan dosya seçilmesini sağlayacak öğelerde bu ekranda bulunmaktadır. Dosya seç butonu tıklandığında bir diyalog kutusu açılmaktadır ve bilgisayardan bir dosya seçilmesi istemektedir.



Şekil 5.21. Seslendirilecek dosyanın seçimi

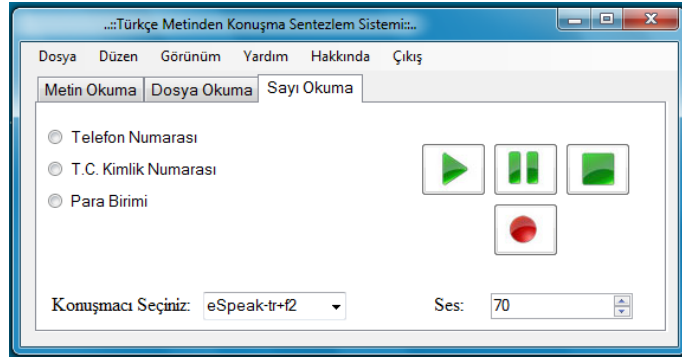
Dosya seçildikten sonra oynat butonuna basıldığında sentezlemede seçilen özelliklere bağlı olarak dosya içerisindeki metin seslendirilmektedir (Şekil 5.22). Seslendirme esnasında açılan formda durdurma ve kayıt işlemini gerçekleştirecek butonlar yer almaktadır. Seslendirme devam ederken bu sesi kaydetme imkânı da bu form sayesinde sunulmaktadır.



Şekil 5.22. Dosyanın seslendirilmesi

### 5.2.5. Sayı Sistemlerinin Seslendirilmesi

Sayı okuma modunda telefon numarası, T.C. kimlik no ve para birimi okuma bölümleri bulunmaktadır (Şekil 5.23). Radyo butonlar seçilen sayı formatı için o formatta metin kutularının çıkmasını sağlamaktadır.



Şekil 5.23. Sayı okuma modu

Telefon numaraları okunurken düz sayı şeklinde okunsa da o numaranın akılda kalması güçleşir. Telefon numaraları alan koduyla birlikte belirli bir formatta okunmaktadır. Bu format genelde 1'li(#), 3'lü(###), 3'lü(###), 2'li(##), 2'li(##) şeklindedir. Örneğin 0-424-233-15-23 şeklindedir. Programda ise bunun için formatlı metin kutusuyla birlikte rakamları bu formata uygun şekilde ayırabilecek bir alt program kullanılmıştır (Şekil 5.24).



Şekil 5.24. Telefon numarası seslendirme

Kimlik numaraları hem akılda tutulurken hem de söylenirken belirli bir kodlamayla söylenebilir. Bunlardan bir tanesi de 3'lü(####), 3'lü(####), 2'li(##), 3'lü(###) tipindedir (Şekil 5.25). Örneğin bir TC kimlik numarası şu şekilde 418-901-25-170 okunabilir.



Şekil 5.25. T.C. kimlik noseslendirme

Para birimi için ise 3'lü sayı sistemlerinin yanında en son hane 2 hanelidir. 3'lü bölümler para biriminin Lira, Dolar gibi olan bölümünü 2'li bölüm ise kuruş ve sent bölümlerini temsil etmektedir.

## 6.SONUÇ

Türkçe Metinden Konuşma Sentezleme sistemlerinde genel olarak izlenen yol; Türkçe dilinin yapısından dolayı yani eklemeli bir dil olduğundan dolayı sentezleme işlemi için en uygun yol eklemeli yöntemlerdir. Benzer çalışmalar incelendiğinde genelde bu yöntemin kullanıldığı görülmüştür. Yine bu çalışmalardan bazıları eklenecek parçalar olarak ikili, üçlü fonemleri kullanmıştır. Genel olarak literatür incelendiğinde Türkçenin en küçük yapı taşının hece olduğu ve heceler yardımıyla bir kelimedenden çokça kelimeler üretildiği de bilinmektedir. O yüzden birleştirilecek seslerin hecelerden seçilmesi daha uygun olmaktadır.

Bir metnin hecelere ayrılmasını sağlamak için çeşitli algoritmalar vardır. Bu algoritmalar yardımıyla metinler hecelere ayrılabilir. Girilen metin sadece yazılardan oluşmadığı zaman ise uygun seslendirme için bazı metin ön işlemleri gerekmektedir. Örneğin kısaltmaların tam okunuşu, sayısal verilerin para için farklı, TC kimlik no için farklı ve telefon numaraları için farklı durumlarda farklı okunması gibidir. Ayrıca "+", "-", "/" sembollerinin pozitif, negatif veya taksim şeklinde seslendirilmesi de metin ön işlem sürecinde yapılması gerekenlerdir. Ön işlemden geçerek hecelere ayrılan metin bir sonraki aşamada eklemeli yöntem kullanılarak birleştirilmektedir.

Eklemeli yöntemlerde karşılaşılan en büyük zorluk farklı frekans ve tonlamalarda kaydedilen ses parçalarının birleştirildiği zaman gerçek insan sesine benzerliğinin düşük olmasıdır. Bunun için önceden kaydedilen ses parçaları genellikle Hanning penceresi olmak üzere bir pencere fonksiyonundan geçirilerek veritabanına aynı frekanslarda eklenmesi gerekmektedir. Bu sayede ses parçaları aynı frekanslarda oluşturulabilir. Bir sonraki aşamada ise bu parçaların birleştirilirken geçişlerinin sağlanması için uygun bir birleştirme yönteminin kullanılmasıdır. OLA ve türevleri olan SOLA, TD-PSOLA, WSOLA fonksiyonları bu birleştirme işlemi için kullanılmaktadır. Bu yöntemlerden en çok kullanılanı ise Zaman-Alan Senkronize Ton Örtüşme ve Ekleme Algoritmasıdır (TD-PSOLA). Bu algoritmanın en büyük avantajı süre değiştirme, perde frekansı modifikasyonu gibi işlemleri eş zamanlı olarak yapabilesidir.

TD-PSOLA kullanılacak eklemeli bir MKS sisteminde kelimeyi oluşturan ses parçaları birbirine faz uyumlu bir şekilde birleştirilebilmektedir. Oluşturulacak sistemin kullanıcıyla etkileşimin maksimum seviyede olabilmesi için kullanışlı bir ara yüz hazırlanması gerekmektedir.

Yapılan birinci çalışmada Türkçe metinler için Eklemeli (Concatenative) Sentezleme yöntemi kullanılarak bir MKS sistemi tasarlanmıştır. Eklenerek parça olarak heceler seçilmiştir. Üç heceli örnek bir kelime ele alınarak tasarlanan sistem Matlab alt yapısıyla programlanmıştır. Hazırlanan sistem de önceden kaydedilen hecelerın ötümsüz yani ses boşlukları silinmiştir. Bunun için bir alt program hazırlanıp heceler bu programdan geçirilmiştir. Bir sonraki aşamada ise birleştirilecek hecelerın faz ve frekans uyumlarının oluşturulabilmesi için bir ses işleme alt programı tasarlanmıştır. Burada bir filtreleme algoritması kullanılmıştır. Bu işlem için Hanning Penceresi kullanılmıştır. Yapılan literatür araştırmasında ise eklemeli sistemlerde en çok kullanılan filtrenin Hanning Penceresi olduğu görülmüştür. Ön işlemde geçirilerek tekrar kaydedilen heceleri birleştirmek için PSOLA algoritması kullanılmıştır. Bu algoritma Ton Örtüşme Ekleme algoritması olarak bilinmektedir. Birleştirilecek hecelerın belirli kısımlarının üst üste gelmesi mantığına dayanan algoritmada bir önceki hecenin son kısmına fade out (yavaş yavaş kaybolma) bir sonraki hecenin ilk kısmına ise fade in (yavaş yavaş belirme) efekti verilmesi mantığına dayanır. Bu işlemler sonucu oluşan kelimenin anlaşılabilir bir seviyede olduğu görülmüştür.

Hazırlanan ikinci uygulamada ise Microsoft SAPI Kütüphanesi yardımıyla bir konuşma sentezleme sistemi oluşturulmuştur. Veritabanı olarak Mbrola projesiyle oluşturulan Türkçe erkek ve kadın sesleri kullanılmıştır. Fonem seslendirmesi mantığına dayanan sistemde noktalama işaretlerinin, rakamsal verilerin ve kısaltmaların başarılı bir şekilde seslendirildiği görülmüştür. C# dili ile hazırlanan sistem de kullanıcı arayüzü hazırlanan ilk sisteme göre daha kullanışlıdır. Ses düzeyinin anlaşılabilirliği yeterli olsa da daha robotik bir ses sentezlenmiştir.

Uygulamalar bu haliyle bir MKS sisteminden beklenen tüm özellikleri göstermese de yeterli seviyede konuşma sentezleme görevini gerçekleştirmektedir. Uygulamaların daha verimli olabilmesi için bu iki sistemin birleştirilmesi veya en azından iki uygulamanın işlevlerini tek başına yapabilecek bir sistemin oluşturulması gerekmektedir. Ayrıca Türkçe seslendirme kurallarının sisteme tam olarak yerleştirilmesi gerekmektedir. Ulama, ünsüz sertleşmesi ve yumuşaması gibi Türkçe diline özgü kuralların sisteme dahil edilmesi

gerekmektedir. Her geen gn yeni kısaltmaların dilimize eklendiđini varsayarsak kısaltma veritabanının srekli gncel tutulması gerekmektedir. Bunun yanında metin ierisindeki rakamların ne tr rakamlar olduđunun anlařılması iin alt programlar eklenmelidir. Vurgu ve tonlamada ki eksiklikler de giderilmesi gereken sorunlar arasındadır. Bu konuda literatrde henz tam performanslı bir alıřma bulunmamakla beraber MKS sistemlerinde giderilmesi gereken bir durumdur. Bunun yanında daha iyi bir ses kalitesi iin veritabanının da gncellenmesi gerekmektedir. Web tabanlı uygulama da geliřtirilirse internet kullanıcılarının devamlı eriřebileceđi cretsiz bir MKS sistemi oluřturulabilir. Daha sonraki alıřmalarda bu tr bir MKS sistemi oluřturulmaya alıřacaktır.



## KAYNAKLAR

- [1] **Eker, B.**, 2002. Turkish Text to Speech System, *Yüksek Lisans Tezi*, Bilkent Üniversitesi, Mühendislik ve Fen Bilimleri Enstitüsü, Ankara
- [2] **Yılmaz, A.E.**, 2009. Türkçe Metinden Konuşma Sentezleme Uygulamaları için bir Veri Sözlük Seti ve Yazılım Çerçevesi Önerisi. Elektronik Mühendisliği Bölümü, Ankara Üniversitesi. IEEE 17. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU-2009), Side/Antalya, Türkiye.
- [3] **Uslu İ.B.**, 2010/2. Metinden Konuşma Sentezleme, TMMOB Elektrik Mühendisleri Odası *Ankara Şubesi Haber Bülteni*, 2, 12-16
- [4] **Aşlıyan R., Günel K., ve Filiz A.**, “Türkçe Otomatik Heceleme Sistemi ve Hece İstatistikleri”, *Akademik Bilişim 2006 BilgiTek IV*, 9-11 Şubat 2006, Pamukkale Üniversitesi, Denizli.
- [5] **Şentürk T. ve Adalı E.**, 2010. Türkçe Metin Seslendirme, *BBM Dergisi*, 4, 35-51
- [6] **Can B.**, 2007. Bir hece-tabanlı Türkçe sesli ifade tanıma sisteminin tasarımı ve gerçekleştirimi, *Yüksek Lisans Tezi*, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara
- [7] **Verhelst, W. and Roelands, M.**, An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech, Proceedings for the International Conference of Acoustics, Speech, and Signal Processing, Minneapolis 1993.
- [8] **Görmez, Z.**, 2009. Implementation of a text-to-speech system with machine learning algorithms in Turkish, *Yüksek Lisans Tezi*, Fatih Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [9] **Flanagan, J.**, Speech Analysis, Synthesis, and Perception, Springer-Verlag, Berlin-Heidelberg-New York, 1972
- [10] **Schroeder M.**, A Brief History of Synthetic Speech, Speech Communication vol.13, pp. 231-237, 1993
- [11] **Güldah, K.**, 2009. Türkçe Metin Seslendirme, *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul

- [12] **Klatt D.**, Review of Text-to-Speech Conversion for English, *Journal of the Acoustical Society of America*, JASA vol. 82(3): 737-793, 1987
- [13] **Erdemir, C.**, 2010. Türkçe metin seslendirme için doğal konuşma sentezleme, *Yüksek Lisans Tezi*, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul
- [14] **Allen J.**, Hunnicutt S., Klatt D., From Text to Speech: The MITalk System, Cambridge University Press, 1987
- [15] **Ünalı, İ.**, 2007. Taşınabilir cihazlar için Türkçe metinden konuşma sentezleme sistemi, *Yüksek Lisans Tezi*, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- [16] **Ergenç, İ.**, Spoken Language and Dictionary of Turkish Articulation, Multilingual Yabancı Dil Yayınları, İstanbul, 2002.
- [17] **Lemmetty, S.**, (1999), "Review of speech synthesis technology" *Master Thesis*, Helsinki University of Technology, March 1999.
- [18] **Verhelst, W.**, "Overlap-Add Methods for Time-Scaling of Speech", *Speech Communication*, vol. 30, nr. 4, 2000.
- [19] **Oskay, B.**, 2002, "Automatic Modelling of Turkish Prosody", *Yüksek Lisans Tezi*, ODTÜ, Fen Bilimleri Enstitüsü, Ankara
- [20] [http://tcts.fpms.ac.be/synthesis/introtts\\_old.html](http://tcts.fpms.ac.be/synthesis/introtts_old.html) Dutoit T., A Short Introduction to Text-to-Speech Synthesis, Erişim Tarihi: 02.04.2013
- [21] <http://www.acapela-group.com>, Text to Speech and Voice Solutions, Erişim Tarihi:10.04.2013
- [22] <http://www.nuance.com/realspeak>, Erişim Tarihi: 01.04.2013
- [23] <http://tcts.fpms.ac.be/synthesis/mbrola.html>, Mbrola Project Homepage, Erişim Tarihi:01.04.2013
- [24] [www.dilimiz.gen.tr](http://www.dilimiz.gen.tr), Türk Dili Kullanımı. Erişim Tarihi: 01 Haziran 2012.
- [25] **Canal Ş.M., Kurnaz S. ve Yılmaz A.E.**, 2010. Türkçe Metinden Konuşma Sentezlemede Yaşanan Sıkıntılar ve Çözüm Yöntemleri, *Havacılık Ve Uzay Teknolojileri Dergisi*, 3, 47-55
- [26] **Artuner, H.**, 1994. Bir Türkçe Fonem Kümeleme Sistemi Tasarımı ve gerçekleştirimi, *Doktora Tezi*. Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- [27] **Parson, T.**, 1986. Voice and Speech Processing, McGraw-Hill.

- [28] **Ergenç, İ.**, 1989, Türkiye Türkçesinin Görevsel Sesbilimi, Engin Yayınları.
- [29] **Demircan, Ö.**, Türkiye Türkçesinin Ses Düzeni, Türkiye Türkçesindeki Sesler, Türk Dil Kurumu Yayınları, Ankara, 1979.
- [30] **User, H. Ş.**, Başlangıcından Günümüze Türk Yazı Sistemleri, Akçağ Yayınları, Ankara, 2006.
- [31] **Salor, Ö., Pellom, B. L., Ciloglu, T. ve Demirekler, M.**, "Turkish Speech Corpora and Recognition Tools Developed by Porting SONIC: Towards Multilingual Speech Recognition", *Computer Speech and Language*, Cilt 21, 580- 593, 2007.
- [32] **Zülfikar, H.**, Doğru Yazalım, Doğru Okuyalım, Türk Dil Kurumu - Rapor, Ağustos 1999.
- [33] **Uslu İ.B.**, 2010/3. Türkçe Metinden Konuşma Sentezlemede Bugünkü Durum, TMMOB Elektrik Mühendisleri Odası Ankara Şubesi Haber Bülteni, 2, 12-16
- [34] Türkçe İmla Kılavuzu - Türk Dil Kurumu, 2000
- [35] **Çakır E. R., Yıldırım T.**, 2009. Rf tabanlı bilgilendirme sistemi. İstanbul.
- [36] <http://ccrma.stanford.edu/courses/422/projects/WaveFormat>, Wave Dosya Formatı 01.04.2009
- [37] <http://tr.wikipedia.org/wiki/wav>, wav, Erişim Tarihi: 10 Mart 2013
- [38] <http://www.atasoyweb.net/Spektral-Sizinti-Ve-Pencere-Fonksiyonlari>, Sizinti ve pencere fonksiyonlari, 10 Mart 2013
- [39] **Makhoul, J. and El-Jaroudi, J.**, "Timescale modification in medium to low rate coding", *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 1705-1708, IEEE, 1986.
- [40] **Moulines, E. and F. Charpentier**, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones", *Speech Communication* 9, pp.453-467, 1990.
- [41] **Moulines, E. and Verhelst, W.**, "Time-domain and frequency-domain techniques for prosodic modification of speech" in kleijn and Paliwal(eds.), *Speech Coding and Synthesis*, pp. 519-555, Elsevier Science B.V., Natherlands, 1995.
- [42] **Sanjaume, J. B.**, Audio Time-Scale Modification in the Context of Professional Audio Post-production, Phd. Thesis, Universitat Pompeu Fabra, Barcelona, 2002.

- [43] **Oflazer, K. ve Inkelas, S.**, “The Architecture and the Implementation of a Finite State Pronunciation Lexicon for Turkish”, *Computer Speech and Language*, Cilt 20, No 1, 80-106, 2006.

## ÖZGEÇMİŞ

1987 yılında Elazığ'da doğdu. İlk, orta ve lise öğrenimini Elazığ'da tamamladıktan sonra 2003 yılında Fırat Üniversitesi, Teknik Eğitim Fakültesi, Elektronik Bilgisayar Eğitimi Bölümü, Bilgisayar Öğretmenliğini kazandı. 2007 yılında bu bölümden mezun oldu. Yine 2010 yılında Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elektronik Bilgisayar Eğitimi Bölümü, Telekomünikasyon Eğitimi Anabilim Dalında yüksek lisans eğitimine hak kazandı. 2007 yılında Milli Eğitim Bakanlığında Bilişim Teknolojileri Teknik Öğretmeni olarak göreve başladı. Halen Malatya Akçadağ Şehit Alper Güde Teknik Lise ve Çok Programlı Lisesinde Bilişim Teknolojileri Öğretmeni olarak görev yapmaktadır. Yabancı dili İngilizce olup 2012 yılında evlenmiştir.

ilhamisel@msn.com