

**ANKARA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**DOKTORA TEZİ**

**KONUŞMA İŞLEME VE TÜRKÇENİN DİLBİLİMSEL ÖZELLİKLERİNİ  
KULLANARAK METİNDEN DOĞAL KONUŞMA SENTEZLEME**

**İbrahim Baran USLU**

**ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI**

**ANKARA**

**2012**

**Her hakkı saklıdır**

## TEZ ONAYI

İbrahim Baran USLU tarafından hazırlanan “**Konuşma İşleme ve Türkçenin Dilbilimsel Özelliklerini Kullanarak Metinden Doğal Konuşma Sentezleme**” adlı tez çalışması 16.05.2012 tarihinde aşağıdaki jüri tarafından oy birliği ile Ankara Üniversitesi Elektronik Mühendisliği Anabilim Dalı’nda **DOKTORA TEZİ** olarak kabul edilmiştir.

**Danışman :** *Prof.Dr. Hakkı Gökhan İLK*

**Jüri Üyeleri :**

*Prof.Dr. Hakkı Gökhan İLK*  
*Ankara Üniversitesi Elektronik Mühendisliği Anabilim Dalı*

*Doç.Dr. Ziya TELATAR*  
*Ankara Üniversitesi Elektronik Mühendisliği Anabilim Dalı*

*Doç.Dr. Sıtkı Çağdaş İNAM*  
*Başkent Üniversitesi Elektrik-Elektronik Mühendisliği Anabilim Dalı*

*Yrd.Doç.Dr. Semih BİNGÖL*  
*Hacettepe Üniversitesi Elektrik-Elektronik Mühendisliği Anabilim Dalı*

*Yrd.Doç.Dr. Fikret ARI*  
*Ankara Üniversitesi Elektronik Mühendisliği Anabilim Dalı*

**Yukarıdaki sonucu onaylarım**

**Prof.Dr. Özer KOLSARICI**  
**Enstitü Müdürü**

## ÖZET

Doktora Tezi

### KONUŞMA İŞLEME VE TÜRKÇENİN DİLBİLİMSEL ÖZELLİKLERİNİ KULLANARAK METİNDEN DOĞAL KONUŞMA SENTEZLEME

İbrahim Baran USLU

Ankara Üniversitesi

Fen Bilimleri Enstitüsü

Elektronik Mühendisliği Anabilim Dalı

Danışman: Prof.Dr. Hakkı Gökhan İLK

Metinden konuşma sentezleme; yazının geliştirilen bir sistem tarafından otomatik olarak okutulmasıdır. Sentezlenen konuşmaya doğal konuşmadaki özelliklerin kazandırılması, bu tezin amacını oluşturmaktadır. Tezde, yöntem olarak eklemeli sentezleme tercih edilmiştir. Konuşma; kayıtlı ikili sesbirimlerin (difonların) uygun şekilde örtüştürülüp eklenmesiyle sentezlenmektedir. Birleştirme öncesi, ikili sesbirimlerin süre, temel frekans ve enerjileri uyumlandırılmaktadır.

Bu tezde Türkçe metinden konuşma sentezleme için yeni bir ezgi modeli geliştirilmiştir. Bu modelde, öncelikle süre için; yazılı Türkçede en çok rastlanan digram grupları dikkate alınarak hareket edilmiştir. Digramlardan karşılık gelen difonlara geçilerek ortalama difon süreleri ile ilk sentez yapılmıştır. Daha sonra sentezlenen sözcüklere difon sayısına göre belirlenen katsayılarla süre değiştirme işlemi uygulanmıştır.

Ezgi modelimizde vurgu için, çekimli fiiller ve tamlamalar üzerinde çalışılmıştır. Olumlu, olumsuz, olumlu-soru ve olumsuz-soru formlarında fiil çekimleri için vurguyu en doğal şekilde veren sinyal işleme yöntemleri (süre, perde frekansı ve enerji değişimleri) sistematik bir şekilde incelenmiştir. Seçilen cümlelerde tamlamalar ve söz grupları üzerinde durularak öbek vurgusunun yer aldığı bölüm için kurallar belirlenmiştir. Ayrıca bazı noktalama işaretlerinin ortalama durak süreleri belirlenmiş ve modele eklenmiştir.

Geliştirilen modelin başarımı öznel dinleme testleri ile ölçülmüştür (CMOS: Comparative Mean Opinion Score).

Sonuç olarak; konuşma işleme ve Türkçenin dilbilimsel özellikleri kullanılarak, Türkçe metinden doğal konuşma sentezleme için, çekimli fiiller, tamlamalar ve bazı noktalama işaretleri üzerinde odaklanan bir ezgi modeli önerilmiştir. Önerilen bu modelle, sentezlenen konuşmanın dinleyiciler tarafından CMOS skoruna göre ortalama 1,45/5,00 puan daha doğal bulunduğu sonucu elde edilmiştir.

**Mayıs 2012, 90 sayfa**

**Anahtar Kelimeler:** Metinden konuşma sentezleme, difon, ezgi modeli, çekimli fiiller, CMOS testi, doğal dil işleme.

## **ABSTRACT**

Ph.D. Thesis

### **SYNTHESIZING NATURAL SPEECH FROM TEXT USING SPEECH PROCESSING AND LINGUISTIC PROPERTIES OF TURKISH**

İbrahim Baran USLU

Ankara University

Graduate School of Natural and Applied Sciences

Department of Electronics Engineering

Supervisor: Prof.Dr. Hakkı Gökhan İLK

Text-to-speech synthesis is the automatic reading of a text by a system. The aim of this thesis is to come up with synthesized speech which carries the properties of natural speech. Concatenative synthesis was chosen as the method and speech was synthesized by diphone concatenations. Before combining the diphones, proper duration, pitch frequency and energy modifications were performed.

A novel intonation model for Turkish text-to-speech synthesis is implemented in this thesis. First, for the duration modelling part, the most frequent digrams in written Turkish were determined and this was the starting point. Using the digrams mean diphone durations were calculated and used for the initial synthesis. Then while synthesizing words, proper duration modifications were applied according to the number of diphones in the words.

For the stress part of our intonation model, verb inflections and compounds were examined. Affirmative, negative, positive and negative interrogative verbs were studied systematically in terms of signal processing techniques (duration, pitch frequency and energy modifications) for the most natural sounding intonation. Rules were determined for the compound stress part in the selected sentences. Also mean durations of some punctuation marks were determined and added to the model.

The developed model was evaluated by CMOS (Comparative Mean Opinion Score) tests.

As a result, a novel intonation model for synthesizing natural speech from text using speech processing and linguistic properties of Turkish is proposed in this thesis. The model focuses on verb inflections, phrases and some punctuation marks. The results show that our model provides an improvement of 1.45/5.00 points in average, in the CMOS scores.

**May 2012, 90 pages**

**Key Words:** Text-to-speech synthesis, diphone, intonation model, inflected verbs, CMOS test, natural language processing.

## TEŞEKKÜR

Öncelikle çalışmalarımı izleyen ve süreç boyunca fikirleriyle tezime yön veren danışmanım sayın Prof.Dr. Hakkı Gökhan İLK'e (Ankara Üniversitesi) teşekkür ederim. Emekleri ve desteği için kendisine şükranlarımı sunarım.

Tez izleme komitemde yer alan ve tezime değerli katkılarda bulunan hocalarım Doç.Dr. Ziya TELATAR'a (Ankara Üniversitesi) ve Yrd.Doç.Dr. Semih BİNGÖL'e (Hacettepe Üniversitesi); tez jürimde yer alıp tezimi okuyan ve tezime önemli katkılarda bulunan sayın Doç.Dr. Sıtkı Çağdaş İNAM'a (Başkent Üniversitesi) ve Yrd.Doç.Dr. Fikret ARI'ya (Ankara Üniversitesi), verdiği fikirlerle tezime çok önemli katkılar sağlayan sayın Doç.Dr. Asım Egemen YILMAZ'a (Ankara Üniversitesi) içten teşekkür ederim.

Difon kayıtlarını tezde kullanımımıza sunarak çalışmalarımıza sesini veren sayın Doç.Dr. Özgül SALOR'a (Gazi Üniversitesi) teşekkürü borç bilirim. Ayrıca, yaptıkları yardımlardan ötürü Yrd.Doç.Dr. Barış BOZKURT'a (İzmir Yüksek Teknoloji Enstitüsü) ve Dr. Oytun TÜRK'e (DFKI GmbH, German Research Center for Artificial Intelligence) de teşekkür ederim.

Karşılaşılan zorlukların aşılmasında yardımlarını esirgemeyen arkadaşlarım Faruk KURAL'a, Utku YARGIÇOĞLU'na, Levent ÖZPARLAK'a ve Mehmet KORKMAZ'a teşekkürlerimi sunarım.

Dilbilim ve Türk Dili konularında yardımcı olan sayın Prof.Dr. İclâl ERGENÇ (Ankara Üniversitesi) ve sayın Prof.Dr. Nurettin DEMİR (Başkent Üniversitesi) hocalarıma teşekkür ederim.

Aynı yerde çalışmaktan onur duyduğum sayın Yusuf Tansel İÇ'e ve Onur KOÇAK'a verdikleri destekten ötürü teşekkür ederim.

Ankara Üniversitesi Fen Bilimleri Enstitüsünün tüm çalışanlarına, başta önceki Enstitü Müdürlerinden sayın Prof. Dr. Orhan ATAKOL hocamız olmak üzere, teşekkürü borç bilirim.

Bu zorlu süreçte her zaman yanımda olan aileme şükranlarımı sunarım.

Bu tez; beni yetiştiren tüm hocalarıma ve aileme ithaf edilmiştir.

İbrahim Baran USLU  
Ankara, Mayıs 2012

## İÇİNDEKİLER

|                                                        |     |
|--------------------------------------------------------|-----|
| ÖZET .....                                             | i   |
| ABSTRACT .....                                         | ii  |
| TEŞEKKÜR .....                                         | iii |
| SİMGELER VE KISALTMALAR DİZİNİ .....                   | vi  |
| ŞEKİLLER DİZİNİ .....                                  | vii |
| ÇİZELGELER DİZİNİ .....                                | ix  |
| 1. GİRİŞ .....                                         | 1   |
| 1.1 Problemin Tanımı .....                             | 2   |
| 1.2 Çalışmanın Amacı .....                             | 2   |
| 1.3 Dilbilim ve Doğal Dil İşleme .....                 | 3   |
| 1.4 MKS Sistemlerinin Kullanım Alanları .....          | 4   |
| 1.5 Tarihsel Gelişim .....                             | 4   |
| 1.6 Türkçe İçin Ezgi Modelleme .....                   | 8   |
| 1.7 Özgünlük İddiamız .....                            | 9   |
| 2. KURAMSAL TEMELLER .....                             | 10  |
| 2.1 Konuşma Sinyalinin Özellikleri .....               | 10  |
| 2.1.1 Ünlüler .....                                    | 11  |
| 2.1.2 Ünsüzler .....                                   | 11  |
| 2.2 MKS Yöntemleri .....                               | 12  |
| 2.2.1 Kural tabanlı sentezleyiciler .....              | 16  |
| 2.2.2 Söyleyiş sentezleyicileri .....                  | 18  |
| 2.2.3 Eklemeli sentezleyiciler .....                   | 19  |
| 2.3 Türkçe MKS Sistemleri İçin Literatür Tarama .....  | 19  |
| 3. MATERYAL VE YÖNTEM .....                            | 23  |
| 3.1 Difon Veritabanı .....                             | 23  |
| 3.2 MKS Sistemleri İçin Sinyal İşleme Yöntemleri ..... | 32  |
| 3.2.1 Sinüsoidal modelleme .....                       | 32  |
| 3.2.2 HNM (Harmonic/ Noise Model) .....                | 32  |
| 3.2.3 TD-PSOLA .....                                   | 33  |
| 3.2.4 LP-PSOLA .....                                   | 33  |
| 3.2.5 MBROLA .....                                     | 33  |
| 3.3 TD-PSOLA Yöntemi .....                             | 34  |
| 3.3.1 Perde işaretlerinin bulunması .....              | 37  |
| 3.3.2 Süre değiştirme .....                            | 39  |
| 3.3.3 Temel frekans değiştirme .....                   | 42  |
| 3.3.4 Enerji değiştirme .....                          | 43  |
| 4. DOĞAL KONUŞMA İÇİN EZGİ MODELLEME .....             | 45  |
| 4.1 MKS Sistemleri İçin Ezgi Modelleri .....           | 48  |
| 4.2 Geliştirilen Süre Modeli .....                     | 49  |
| 4.3 Fujisaki Ezgi Modeli .....                         | 50  |
| 4.4 Geliştirilen Ezgi Modeli .....                     | 55  |
| 4.4.1 Çekimli fiillerde vurgu .....                    | 55  |
| 4.4.2 Sözcük ve öbek vurgusu .....                     | 60  |
| 4.4.3 Noktalama işaretleri .....                       | 64  |
| 5. BULGULAR .....                                      | 66  |

|                                                              |           |
|--------------------------------------------------------------|-----------|
| <b>6. SONUÇLAR VE TARTIŞMA.....</b>                          | <b>71</b> |
| <b>6.1 Tartışma.....</b>                                     | <b>73</b> |
| <b>6.2 Gelecek Çalışmalar .....</b>                          | <b>73</b> |
| <b>KAYNAKLAR.....</b>                                        | <b>75</b> |
| <b>EK 1 Türkçe Metinlerin İstatistiksel Özellikleri.....</b> | <b>77</b> |
| <b>EK 2 Diğer Dillerde Yapılan MKS Çalışmaları.....</b>      | <b>82</b> |
| <b>ÖZGEÇMİŞ.....</b>                                         | <b>89</b> |

## SİMGELER VE KISALTMALAR DİZİNİ

|            |                                                     |
|------------|-----------------------------------------------------|
| $\alpha_d$ | Difon süre deęiřtirme parametresi                   |
| CMOS       | Comparative Mean Opinion Score                      |
| DDİ        | Doęal Dil İřleme                                    |
| $F_0$      | Temel frekans                                       |
| HNM        | Harmonic plus Noise Model                           |
| LP-PSOLA   | Linear Prediction Pitch Synchronous OverLap and Add |
| LPC        | Linear Predictive Coding                            |
| MBROLA     | Multi Band Resynthesis OverLap and Add              |
| MKS        | Metinden Konuřma Sentezleme                         |
| NLP        | Natural Language Processing                         |
| SMM        | Saklı Markov Modeli                                 |
| TD-PSOLA   | Time Domain Pitch Synchronous OverLap and Add       |
| TEO        | Teager Enerji Operatörü                             |
| TTS        | Text-To-Speech                                      |

## ŞEKİLLER DİZİNİ

|            |                                                                                          |    |
|------------|------------------------------------------------------------------------------------------|----|
| Şekil 1.1  | MKS sistemlerinin tarihsel gelişimi.....                                                 | 5  |
| Şekil 1.2  | Kratzenstein'in çınlayıcıları.....                                                       | 5  |
| Şekil 1.3  | VODER konuşma sentezleyicisi (Klatt 1987).....                                           | 6  |
| Şekil 1.4  | Konuşmanın kaynak-süzgeç oluşum modeli.....                                              | 7  |
| Şekil 2.1  | “n1s ” difonu dalgaşekli, “n1” fonemi ötümlü iken,<br>“s” fonemi ötümsüzdür.....         | 10 |
| Şekil 2.2  | LPC analiz süzgeci.....                                                                  | 13 |
| Şekil 2.3  | Formant frekanslarının LPC analiz filtresinin<br>genlik spektrumundan elde edilmesi..... | 14 |
| Şekil 2.4  | Eklemeli sentezlemede işlemlerin akış şeması.....                                        | 15 |
| Şekil 2.5  | a. Seri formant sentezleyici, b. Paralel formant<br>sentezleyici.....                    | 16 |
| Şekil 2.6  | PARCAS model seri-paralel formant sentezleyici.....                                      | 17 |
| Şekil 2.7  | Konuşma üretim mekanizması ve<br>modellenmesi (Palo 2006).....                           | 18 |
| Şekil 3.1  | Taşıyıcı cümle içerisine kaydedilmiş olan “o d ” difonu.....                             | 24 |
| Şekil 3.2  | Elden geçirilmiş ve ayrıştırılmış olan “o d ”<br>difonu dalgaşekli.....                  | 25 |
| Şekil 3.3  | Matlab GUI'de geliştirilen MKS arayüzü.....                                              | 26 |
| Şekil 3.4  | Geliştirilen arayüzdeki süre ayarı özelliği.....                                         | 27 |
| Şekil 3.5  | Geliştirilen arayüzdeki perde frekansı ayarı özelliği.....                               | 28 |
| Şekil 3.6  | Geliştirilen arayüzdeki enerji ve örtüşme<br>perde sayısı ayarı özelliği.....            | 29 |
| Şekil 3.7  | Geliştirilen arayüzde çalışan işlemlerin akış şeması.....                                | 30 |
| Şekil 3.8  | PSOLA yöntemiyle örtüştürüp ekleme.....                                                  | 34 |
| Şekil 3.9  | PSOLA yöntemiyle “k2a1” + “a1r2” birleşmesi.....                                         | 35 |
| Şekil 3.10 | PSOLA yöntemiyle “o k2” + “k2a1” birleşmesi.....                                         | 35 |
| Şekil 3.11 | PSOLA yöntemiyle bir perde örtüştürüp ekleme sonucu.....                                 | 36 |
| Şekil 3.12 | PSOLA yöntemiyle üç perde örtüştürüp ekleme sonucu.....                                  | 36 |
| Şekil 3.13 | PSOLA yöntemiyle altı perde örtüştürüp ekleme sonucu.....                                | 37 |
| Şekil 3.14 | “m e2” difonunda, TEO ile perde işaretlerinin bulunması.....                             | 39 |
| Şekil 3.15 | “m e2” difonunda, dalgaşekli tepeleri ile<br>perde işaretlerinin bulunması.....          | 39 |
| Şekil 3.16 | “m e2” difonunda, özilinti işlevi ile<br>perde işaretlerinin bulunması.....              | 40 |
| Şekil 3.17 | PSOLA ile süre uzatma.....                                                               | 41 |
| Şekil 3.18 | PSOLA ile süre kısaltma.....                                                             | 41 |
| Şekil 3.19 | “h e2” difonunun perde frekansının %20 artırılması                                       |    |

|            |                                                                                                                            |    |
|------------|----------------------------------------------------------------------------------------------------------------------------|----|
|            | ve yeni perde işaretleri.....                                                                                              | 42 |
| Şekil 3.20 | “h e2” difonunun perde frekansının %20 azaltılması<br>ve yeni perde işaretleri.....                                        | 42 |
| Şekil 3.21 | Enerji eşitlenmemiş durumda sentez.....                                                                                    | 43 |
| Şekil 3.22 | Enerji eşitleme sonrası sentez.....                                                                                        | 44 |
| Şekil 4.1  | Üretimsel ve analitik ezgi modeli<br>yaklaşımı (Mixdorff 1998).....                                                        | 49 |
| Şekil 4.2  | Fujisaki ezgi modelinin blok şeması.....                                                                                   | 51 |
| Şekil 4.3  | (a,b) Kayan ortalamalı süzgeçle (N=7)<br>“kokar” sözü F0 eğrisinin düzgünleştirilmesi.....                                 | 52 |
| Şekil 4.4  | a. Tamlama bileşenlerinin b. Aksan bileşenlerinin<br>yer ve genliklerinin bulunması.....                                   | 52 |
| Şekil 4.5  | “kokar” sözü için modellenen aksan bileşeni.....                                                                           | 53 |
| Şekil 4.6  | Ses kaydı olan “kokar” sözünün,<br>a. dalgaşekli, b. F <sub>0</sub> eğrisi.....                                            | 53 |
| Şekil 4.7  | Geliştirilen MKS sisteminin sentezlediği<br>“kokar” sözünün, a. dalgaşekli, b. F <sub>0</sub> eğrisi.....                  | 54 |
| Şekil 4.8  | Ezgisi Fujisaki modeliyle değiştirilen “kokar” sözünün,<br>a. dalgaşekli, b. F <sub>0</sub> eğrisi.....                    | 54 |
| Şekil 4.9  | “Burayı gezdim mi?” olumlu soru cümlesi<br>(“-dim” sesleminde perde frekansı ve<br>enerji değiştirme uygulanmıştır).....   | 56 |
| Şekil 4.10 | “Burayı gezmedim mi?” olumsuz soru cümlesi<br>(“gez” sesleminde perde frekansı<br>ve enerji değiştirme uygulanmıştır)..... | 56 |
| Şekil 4.11 | “Soruları kavramalısın.” cümlesinin ham sentez<br>hali için arayüzdeki ayarlar.....                                        | 57 |
| Şekil 4.12 | “Soruları kavramalısın.” cümlesinin perde frekansı<br>değiştirilmiş hali için arayüzdeki ayarlar.....                      | 58 |
| Şekil 4.13 | Hazırlanan test sayfasında yer alan açıklamalar.....                                                                       | 59 |
| Şekil 4.14 | Test sayfasından bir görünüm.....                                                                                          | 59 |
| Şekil 4.15 | “Otobüsle uzun yola hiç gitmedim” cümlesi<br>için ara yüzdeki ayarlar.....                                                 | 62 |
| Şekil 4.16 | a. “En hakiki mürşit: ilimdir, fendir.” Cümlesinin<br>ham sentezi.....                                                     | 63 |
|            | b. “En hakiki mürşit: ilimdir, fendir.” Cümlesinin<br>ezgili sentezi.....                                                  | 63 |
|            | c. Orijinal kayıt: “En hakiki mürşit: ilimdir, fendir.”.....                                                               | 63 |
| Şekil 5.1  | Sentezlerin aldığı ortalama CMOS Puanları.....                                                                             | 69 |

## ÇİZELGELER DİZİNİ

|             |                                                                                                                                                                     |    |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Çizelge 2.1 | Türkçe MKS için yapılmış olan lisansüstü çalışmalar.....                                                                                                            | 20 |
| Çizelge 3.1 | Difon listesi.....                                                                                                                                                  | 23 |
| Çizelge 3.2 | Sesbirim belirleme kural ve örnekleri.....                                                                                                                          | 31 |
| Çizelge 4.1 | (Şayli 2002)'den alınan ortalama fonem süreleri<br>ve sıkıştırma oranları.....                                                                                      | 46 |
| Çizelge 4.2 | Difon süresi değiştirme parametresi ( $\alpha_d$ ).....                                                                                                             | 50 |
| Çizelge 4.3 | Ezgi modelimizde uygulanan değişiklikler.....                                                                                                                       | 55 |
| Çizelge 4.4 | Süre ve vurgunun birlikte incelendiği cümleler.....                                                                                                                 | 58 |
| Çizelge 4.5 | Noktalama işaretlerinin hesaplanan ortalama süreleri.....                                                                                                           | 60 |
| Çizelge 4.6 | Noktalama işaretlerinin hesaplanan ortalama süreleri                                                                                                                | 64 |
| Çizelge 5.1 | “Burayı gezdim” cümlesinin olumlu, olumsuz, olumlu soru<br>ve olumsuz soru formları için yapılan sentezlere<br>dinleyicilerin verdiği ortalama puanlar.....         | 66 |
| Çizelge 5.2 | "Soruları kavramalısın" cümlesinin olumlu, olumsuz,<br>olumlu soru ve olumsuz soru formları için yapılan<br>sentezlere dinleyicilerin verdiği ortalama puanlar..... | 67 |
| Çizelge 5.3 | Test cümlelerine verilen ortalama CMOS puanları.....                                                                                                                | 68 |

## 1. GİRİŞ

Metinden konuşma sentezleme; herhangi bir metnin bir sistem tarafından otomatik olarak okutulması, seslendirilmesidir. Sentezlenen konuşmanın insan sesi doğallığında olması bu tezin amacını oluşturmaktadır. Metinden konuşma sentezlemede, sentezlenen konuşmada uygun ezgi bulunmuyorsa, konuşma kulağa doğal gelmez. Bu yüzden konuşmaya ezgi özelliklerinin katılması gerekir. Bu kolay bir problem değildir. Çünkü konuşma sinyalinin akustik ve fonetik özellikleri henüz tam olarak çözilememiştir.

Metinden konuşma sentezleme (MKS); pek çok dilde üzerinde çalışılan önemli bir araştırma konusudur. Bu konunun uygulama alanları arasında: görme ve konuşma engelli insanlarla iletişim kurma (Braille Teknik Ltd. Şti. – <http://www.brailleteknik.com/jaws.html>, son erişim: 08/02/2012), sesli yanıt ve okuma sistemleri (Loquendo S.p.A. a Telecom Italia Group Company – <http://www.loquendo.com/en/demo-center/tts-demo/>, GVZ Ses tanıma ve sentezleme teknolojileri şirketi – <http://www.gvz.com.tr/index.html>, DİKTE Yöndata Bilgisayar Ltd. Şti. – <http://www.dikte.com.tr/konusmatanima.php>, son erişim: 08/02/2012), dil edinimi ve yabancı dil öğretimi (Google translate – <http://translate.google.com>, son erişim: 08/02/2012), dilbilim çalışmaları, insan-makine etkileşimi vb. sayılabilir.

İlk zamanlarda yapılan sistemler oldukça mekanikti ve bilim-kurgu filmlerinin (2001: *A Space Odyssey* gibi) dışında kullanım alanı bulamamıştı. Zamanla teknolojiadaki gelişmeler ve bu alanda yapılan araştırmaların sonucunda MKS sistemlerinin anlaşılabilirliği ve doğallığı arttı. Makine tarafından sentezlenen konuşmanın kalitesi yükseldikçe, MKS sistemleri gündelik hayata daha çok girmeye başladı.

## 1.1 Problemin Tanımı

Metinden konuşma sentezleme; yazının otomatik olarak okutulmasıdır. Sistemin girdisi metin, çıktısı ise konuşmadır. Kısaca MKS (Metinden Konuşma Sentezleme) olarak adlandırılan bu konunun en genel uygulama alanı, kişinin bir metni okumak yerine dinlemesidir. MKS sistemlerinde son yıllarda tercih edilen yöntem: eklemeli sentez yöntemidir (diğer yöntemler Bölüm 2.2’de ayrıntılı olarak ele alınmıştır). Bu yöntemde önceden kaydedilen temel konuşma parçaları, sentez sırasında uygun duruma getirilerek birleştirilir. Farklı zamanlarda kaydedilmiş olan konuşma parçalarının farklı sentezler için bir araya getirilmesi sonucunda parçalar arasında süre, temel frekans ve enerji uyumsuzluklarının olması kaçınılmazdır. Bunların aşılması ve doğal konuşmadaki ezgi değişimlerinin sentezlenecek konuşmaya kazandırılması gerekir. Çünkü bilindiği üzere, ezgi; konuşmanın daha kolay anlaşılması için konuşmacı tarafından konuşmaya eklenen özelliklerdir. Bir soru cümlesinde soru ekinden hemen önce sesin tizleşmesi veya olumsuzluk durumunu belirtmek için enerjide uygulanan fark, buna örnek olarak verilebilir. Tezde ele alınan problemler; bu uyumsuzlukların giderilmesi, doğru vurgunun ve prozodik etkilerin senteze kazandırılması hedefini taşımaktadır.

## 1.2 Çalışmanın Amacı

Bu çalışmanın amacı; metinden sentezlenen Türkçe konuşmanın doğallığının ve kalitesinin artırılmasıdır. Türkçenin tamamında bunu yapmak oldukça zor bir problem olduğu için, çalışma çekimli fiiller, tamlamalar ve bazı noktalama işaretlerinin vurguları üzerinde yoğunlaşmıştır. Türkçede cümleler isim ve fiil cümleleri olmak üzere ikiye ayrılır. Yükleme fiil olanlar fiil cümlesi, yükleme isim olanlar isim cümlesi olarak adlandırılır. Türkçede fiil çekimlerinin kuralları belirlidir (Banguoğlu 2007, Aydemir ve Yılmaz 2010). Bu çalışmada fiil cümlelerinin ezgisini en doğal şekilde senteze verecek sinyal işleme kuralları belirlenmeye çalışılmıştır.

Bir cümlede asıl vurgunun yer aldığı bölümler vardır. Öbek vurgusu adı verilen bu

vurgu cümlede anlamı belirler ve söyleyiş sırasında baskın akustik ögeler taşır. Tezde öbek vurgusu üzerinde önemle durulmuş, bu vurgunun senteze kazandırılması için çalışılmıştır.

Bir yazıda, durakları belirtmek ve anlamı belirlemek için noktalama işaretleri kullanılır. Çalışmamızda virgül, noktalı virgül ve iki nokta üst üste gibi durak belirten noktalama işaretleri dikkate alınmış ve sentezde bu noktalama işaretlerinde ne kadar boşluk bırakılması gerektiği hesaplanmıştır.

### **1.3 Dilbilim ve Doğal Dil İşleme**

Doğal Dil İşleme (DDİ) (NLP: Natural Language Processing); dilin istatistiksel özelliklerini belirleyip bunlardan çeşitli amaçlar için yararlanan alandır. DDİ'nin uygulamaları için; metin türü tanıma, yazar tanıma, sınıflandırmalar, yazılı dökümanların otomatik olarak başka dillere çevrilmesi, soru-cevap makineleri, otomatik konuşma ve komut anlama, konuşma sentezi, konuşma üretme, otomatik metin özetleme, bilgi sağlama gibi birçok örnek verilebilir (Yılmaz 2009), (Nabiyev 2005). Çalışma kapsamında Türkçe metinlerin istatistiksel özellikleri incelenmiştir ve Ek 1'de anlatılmıştır.

Bu tez, sinyal işleme, bilgisayar bilimleri ve dilbilim alanlarının kesişiminde yer almaktadır. Tez kapsamında öncelikle sıfırdan bir sentezleyicinin tasarlanması ve gerçekleştirilmesi hedeflenmiştir. Daha sonra bu sentezleyicinin dilbilim kuralları ışığında doğal sentezler yapmasını sağlayacak yöntemler araştırılmıştır. Bu noktada Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dilbilim Bölümü'nden alınan "Türkçenin Sesbilimsel Görünümleri-2" adlı derste edinilen bilgilerden yararlanılmıştır.

Tez çalışmasında dilbilimsel kurallar önemli bir yere sahiptir. Sözcük vurgusu ve cümle ezgisi; bazı dilbilimciler tarafından daha önce çalışılmış konulardır (Ergenç 2002, Inkelas ve Orgun 2003, Kabak ve Vogel 2005). Türkçede sözcük vurgusu genel olarak

son hecededir ve her sözcükte bir ana vurgu vardır. Ancak vurgu konusunda istisnai durumlar da karşımıza çıkmaktadır.

#### 1.4 MKS Sistemlerinin Kullanım Alanları

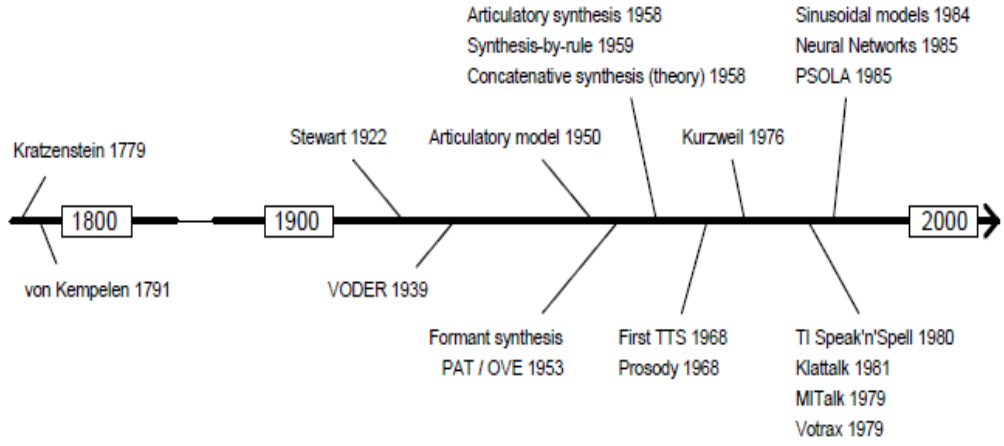
MKS sistemlerinin uygulama alanlarına örnekler:

- Konuşma veya görme engelli kişilerle iletişim
- Sağlık sektörü
- Sesli yanıt sistemleri
- Sesli uyarı sistemleri
- Dilbilim araştırmaları, yabancı dil öğretimi
- Okuma sistemleri
- Konuşan oyuncaklar

Metinden konuşma sentezleyen sistemler, engelli kişiler için çok faydalı olanaklar sağlamaktadır. *Metinden-konuşmaya (TTS: Text-To-Speech)* ve *konuşmadan-metne* dönüştürücüler sayesinde, konuşma veya duyma-engelli kişilerle iletişim kurulabilir. Ayrıca, bir *karakter tanıma sistemi* desteğindeki bir sentezleyici ile görme engelliler için taranan bir metnin okunması da mümkündür. Ek 2’de diğer dillerde yapılan MKS sistemlerine örnekler verilmiştir.

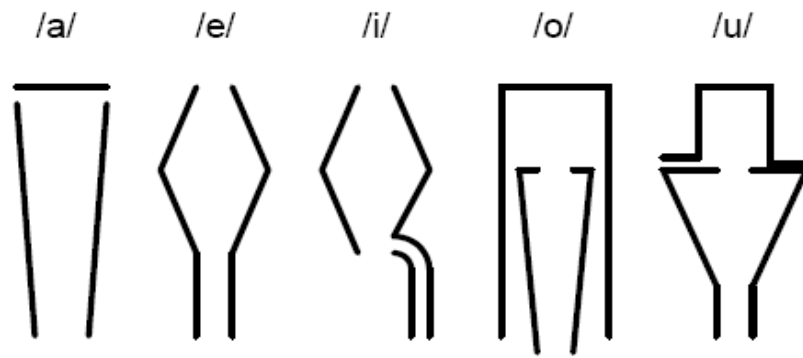
#### 1.5 Tarihsel Gelişim

MKS sistemlerinin tarihsel gelişimi şekil 1.1’de verilmiştir. İlk zamanlar, akustik ve mekanik sistemler üzerinde çalışılmış, elektriğin keşfinden sonra elektriksel üretim sistemleri geliştirilmiştir.



Şekil 1.1 MKS sistemlerinin tarihsel gelişimi

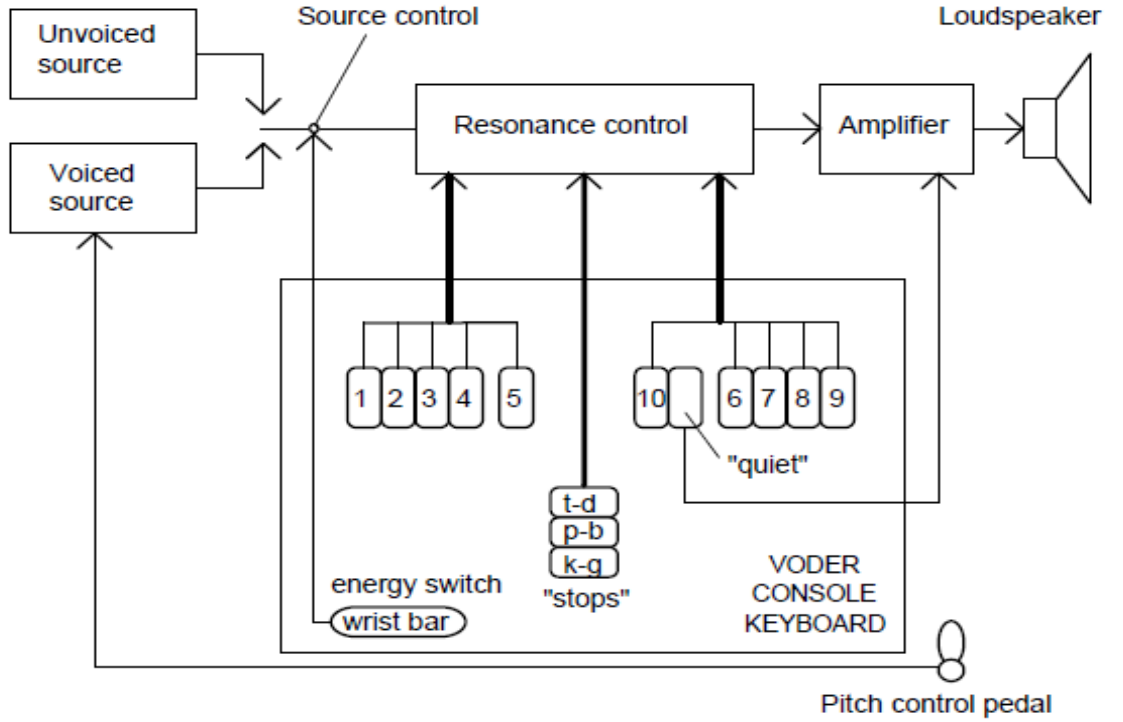
İlk olarak Kratzenstein 1779'da, 'a, e, i, o, u' ünlü sesleri arasındaki fizyolojik farklılıkları belirlemiş ve bu sesleri çınlayıcılar (resonators) yardımıyla yapay olarak sentezlemeyi başarmıştır (şekil 1.2), (Lemmetty 1999).



Şekil 1.2 Kratzenstein'in çınlayıcıları

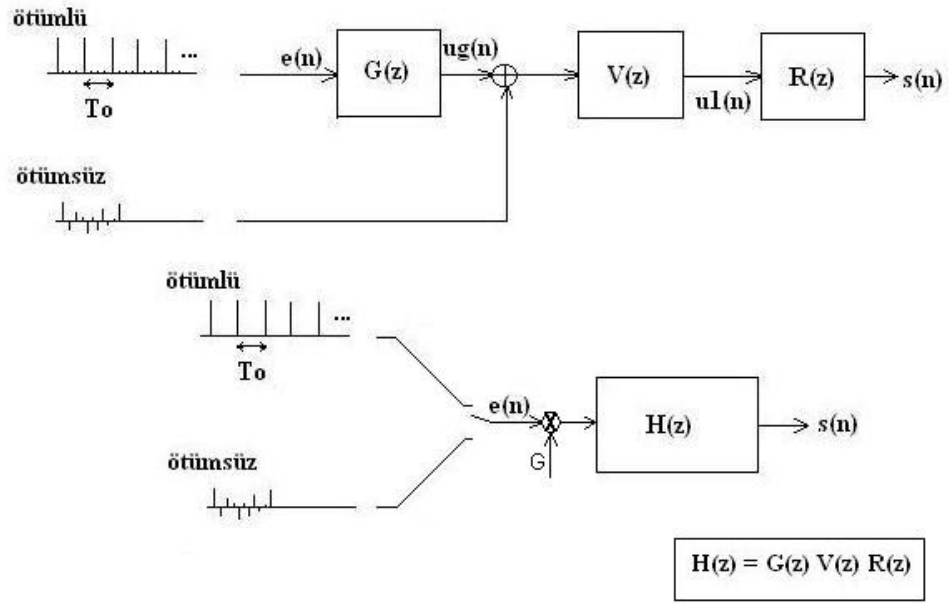
1791'de, Wolfgang von Kempelen akustik telaffuzun ana kaynağının *ses yolu* (*vocal tract*) olduğunu bulmuş; 1838 yılında Willis; ünlüler ile ses yolunun geometrisi

arasındaki bağıntıyı açıklamıştır. 1922’de Stewart tarafından yapılan iki çınlayıcı makine, ünlüleri ilk iki formantlarından sentezlemeyi başarmıştır. Bu sistem ilk elektriksel sentezleyicidir. 1939’da Dudley; ilk konuşma sentezleyici olarak kabul edilen VODER (Voice Operating Demonstrator) adlı sistemi geliştirmiştir.



Şekil 1.3 VODER konuşma sentezleyicisi (Klatt 1987)

El ve ayakla denetlenen bu sistem *kaynak-süzgeç (source-filter)* modeline yakındır. Bu modelde (şekil 1.4),  $G(z)$ ; ses tellerinin,  $V(z)$ ; ağız ve geniz boşluklarının,  $R(z)$  de dudakların etkisini modellemektedir (Özen 2002). Tüm bu filtreleri bir tane eşdeğer  $H(z)$  filtresi ile temsil etmek mümkündür. Bilindiği gibi, ötümlü çerçeveler için girdi sinyaliz;  $T$ : perde periyotlu bir dürtü dizisi iken, ötümsüz çerçeveler için  $e(n)$ ; beyaz gürültüdür.



Şekil 1.4 Konuşmanın kaynak-süzgeç oluşum modeli

Daha sonra 1953'te, Lewrance tarafından PAT (Parametric Artificial Talker) adlı üç çınlayıcıli paralel formant sentezleyici ve Fant tarafından OVE (Orator Verbis Electris) adlı seri formant sentezleyici yapılmıştır.

1958'de Rosen, DAVO (Dynamic Analog of the VOcal Tract) adlı ilk söyleyiş (articulatory) sentezleyicisini gerçekleştirmiştir.

1968'de Umeda tarafından İngilizce için ilk MKS yapılmış, bunu 1976 yılında Kurzweil'in görme engelliler için tasarladığı ilk okuma sistemi izlemiştir. 1979'da, Allen, Hunnicutt ve Klatt (Allen vd. 1987); MITalk adlı MKS sistemini gerçekleştirmişlerdir. Bu sistem ticari amaçla da kullanılmıştır.

Günümüze gelirken, 1984'te MIT'te McAuley ve Quatieri tarafından sinüsoidal modelleme yöntemi önerilmiş, 1985'te Fransız Telekom tarafından PSOLA (Pitch Synchronous OverLap and Add) (Moulines ve Charpentier 1990) yöntemi, örtüşmeli ekleme yöntemi olarak önerilmiştir.

## 1.6 Türkçe İçin Ezgi Modelleme

Türkçe metinden konuşma sentezlemede, konuşmanın doğallığının artırılması için süre ve perde frekansı modelleme üzerinde daha önce bazı çalışmalar yapılmıştır. Abdullahbeşe (2001), frekans konturlarını modellemeye çalışmış, metindeki yapısal özellikleri kullanarak cümle düzeyinde perde frekansı eğrileri oluşturmuştur. Şayli (2002), süre modelleme üzerinde çalışmış ve konuşma parçalarının (fonem ve trifon) ortalama sürelerini hesaplamıştır. Oskay (2002), cümle ezgilerinin otomatik olarak modellenmesi üzerinde çalışmıştır. Vural (2003), sözcük vurgusu ve allofon süreleri üzerine yaptığı çalışmada Festival adlı sentezleyiciyi kullanmıştır. Öztürk (2005), sesbirim sürelerini ve temel frekans konturlarını modellemeye çalıştığı çalışmasında sınıflandırma ve bağlanım ağaçlarını kullanmıştır. Külekçi ve Oflazer (2006), tamlamaların bulunması üzerinde durmuşlar ve bulunan tamlamalara ezgi seviyeleri atamışlardır. Uslu ve İlk (2009), pek çok dilde başarılı sonuçlar veren Fujisaki ezgi modelini bazı Türkçe sözcük ve cümlelere ilk defa uygulamışlardır.

Araştırılan güncel konulardan biri de: duygusal sentezdir. Senteze duygu katmayı amaçlayan bu çalışmalarda hedef; mutlu, üzgün, kızgın, sıkılmış, korkmuş birinin duygularını konuşmaya yansıtmaktır. Bu alanda, Bulut vd. (2005) yaptıkları çalışmada; sesbirim düzeyinde süre, perde frekansı, enerji ve izge (spectrum) değişikliklerinin duygusal senteze olan etkisini incelemişlerdir. Sonuç olarak; izgel zarf değişikliklerinin diğerlerine göre daha etkili, yerel prozodi değişikliklerinde ise; süre değişikliğinin perde frekansı değişikliğinden daha başarılı olduğunu belirtmişlerdir. Burkhardt vd. (2006), farklı dillerde (Türkçe, Yunanca, Almanca ve Fransızca) duygusal sentezler yapıp birbirleriyle karşılaştırmışlardır. Temel frekans, süre ve “jitter” parametreleriyle senteze duygu katmaya çalışmışlar; sonuçta, hem o dile özgü, hem de tüm dillerde ortak noktalar bulunduğunu belirlemişlerdir. Türkçe MKS konusunda yapılan tüm lisansüstü çalışmalar Bölüm 2.3’te ele alınacaktır.

## 1.7 Özgünlük İddiamız

Bu çalışmanın özgün tarafları şunlardır:

- Çekimli fiillerle ilgili vurgu kuralları araştırılmıştır. Olumlu, olumsuz, olumlu soru ve olumsuz soru formundaki fiillerde vurgunun akustik özellikleri kural tabanlı olarak belirlenmiştir.

Bu iddiamızdan aşağıdaki yayın yapılmıştır:

*Uslu, İ.B., Demir, N., İlk, H.G ve Yılmaz, A.E., Bilgisayar bir metni vurgulu okuyabilir mi?, Bilig Türk Dünyası Sosyal Bilimler Dergisi (S-SCI) (2012 yılı, 63. Güz sayısında yayınlanmak üzere kabul edildi)*

- Tamlamalar ve bazı noktalama işaretleri de dikkate alarak sentezlenen cümlelere doğal konuşmadakine yakın ezgi kazandıran bir model önerilmiştir.

Önerilen bu modelden de aşağıdaki yayın yapılmıştır:

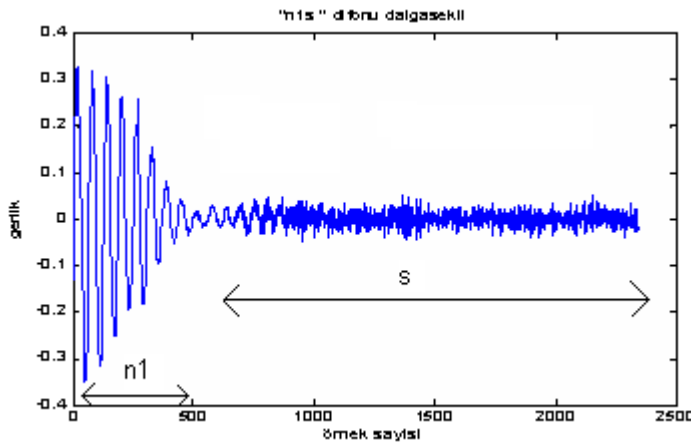
*Uslu, İ.B., İlk, H.G ve Yılmaz, A.E., Türkçe metinden konuşma sentezlemede doğallığın artırılması için öneriler, EMO Bilimsel Dergi, cilt.1, sayı.2, Aralık 2011 s.95-102.*

## 2. KURAMSAL TEMELLER

Bu bölümde, tez çalışmasının temelini oluşturan kuramsal bilgiler verilmektedir. Konuşma sinyalinin ve metinden konuşma sentezleme (MKS) sistemlerinin genel özellikleri ile Türkçe MKS sistemleri için daha önce yapılan çalışmalar bu bölümde incelenmektedir. MKS sistemlerinde üç tür yaklaşım vardır. Bunlar: 1- Kural-tabanlı formant sentezleyiciler, 2- Söyleyiş [articulatory] sentezleyicileri, 3- Eklemeli [concatenative] sentezleyicilerdir. Her üç yaklaşımın da özellikleri aşağıda anlatılmaktadır.

### 2.1 Konuşma Sinyalinin Özellikleri

Konuşma sinyali, 20-30 ms'lik çerçeveler için durağan kabul edilebilir. Konuşma sinyali; neredeyse-periyodik özellikte, enerji değeri büyük, alçak frekanslı bileşenler içeren *ötümlü (sedalı) çerçeveler* ile gürültü benzeri özellik gösteren, enerjisi düşük, yüksek frekanslı bileşenler içeren *ötümsüz (sedasız) çerçeveler*den oluşur. Ünlü harfler (a, e, ı, i, o, ö, u, ü) ve bazı ünsüzler (b, d, g, l, m, n, r), ötümlü [voiced] özellikteyken; patlamalı ünsüzler (p, t) ve bazı sızıcı ünsüzler (h, j, s, ş, z) ötümsüz [unvoiced] özelliktedir (Ergenç 2002). Şekil 2.1'de ötümlü ve ötümsüz çerçeveler içeren bir difon örneği verilmiştir.



Şekil 2.1 “n1s ” difonu dalgaşekli (“n1” fonemi ötümlü iken, “s” fonemi ötümsüzdür)

### 2.1.1 Ünlüler

Oluşumları sırasında ses geçidinde belirli hiç bir takıntıya uğramayan, hiç bir engelle karşılaşmayan seslere ünlü adı verilir (Ergin 1998). Ünlüler; ses yolunun açıklık derecesine, yani çene, dil ve dudakların durumuna göre çeşitlilik gösterir. Ünlüleri çene açısının durumuna, dudakların biçimine ve dilin devinimine göre üç grupta sınıflandırabiliriz (Ergenç 2002):

Çene açısının durumuna göre:

- geniş ünlüler [wide vowels] {a, e, o, ö}
- dar ünlüler [narrow vowels] {ı, i, u, ü}

Dudakların biçimine göre:

- düz ünlüler [unrounded vowels] {ı, i, a, e}
- yuvarlak ünlüler [rounded vowels] {o, ö, u, ü}

Dilin devinimine göre:

- arkadil ünlüleri [back vowels] {ı, a, o, u}
- öndil ünlüleri [front vowels] {i, e, ö, ü}

### 2.1.2 Ünsüzler

Oluşumları sırasında, ses yolunda belirli bir darlık, gevşek veya gergin bir kapantı olur. Seda (ötümlülük) özelliği açısından 3 gruba ayrılabilir (Ergin 1998): 1- ötümsüz karşılıkları olan ötümlü ünsüzler [b, c, d, g, ğ, j, v, z], 2- ötümsüz karşılıkları olmayan ötümlü ünsüzler [l, m, n, r, y], 3- ötümsüz ünsüzler [ç, f, h, k, s, ş, p, t]. Ötümlü sesler; oluşumları sırasında ses tellerinin titrediği seslerdir. Ötümsüz seslerin oluşumları sırasında ses telleri titreşmez. Ünsüzler; çıkış biçimlerine, çıkış yerlerine ve ses tellerinin durumuna göre de sınıflandırılabilir (Ergenç 2002):

Çıkış biçimlerine göre:

1. Patlamalı ünsüzler [plosive consonants]: {b, d, g, p, t, k}
2. Geniz ünsüzleri [nasal consonants]: {m, n}
3. Çarpmalı ünsüzler [rolled]: {r}
4. Yan daralma ünsüzleri [lateral]: {l}
5. Sürtünücü ünsüzler [fricative]: {c, ç, f, h, j, s, ş, v, y, z}

Çıkış yerlerine göre:

1. çift dudak ünsüzleri [bilabial]: {b, p, m}
2. dudak-diş ünsüzleri [labiodental]: {f, v}
3. dilucu-dişardı ünsüzleri [dental]: {d, t}
4. dilucu-dişeti ünsüzleri [palato-alveolar]: {n, r, s, z}
5. dil-öndamak ünsüzleri [alveo-palatal]: {c, ç, j, ş, y}
6. dilucu-öndamak ünsüzleri [apical/ palatal]: {l}
7. dil-artdamak ünsüzleri [velar]: {g, k}
8. gırtlak ünsüzleri [glottal]: {h}

Ses tellerinin titreşimine göre;

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

## 2.2 MKS Yöntemleri

Literatüre bakıldığında, metinden konuşma sentezlemede üç yaklaşım geliştirildiği görülür. Bu yaklaşımlar: 1- Kural-tabanlı formant sentezleyiciler, 2- Söyleyiş sentezleyicileri, 3- Eklemeli sentezleyicilerdir. Birinci grup sentezleyiciler konuşma sinyalinin doğrusal öngörümlü kodlaması [LPC, Linear Predictive Coding] temeline dayanmaktadır. Bu ise; eşitlik (2.1)'de görüldüğü gibi, konuşma sinyalinin n. örneğinin ( $\hat{s}(n)$ ), önceki p adet örneğin doğrusal kombinasyonu şeklinde ifade edilmesidir Burada p; süzgecin derecesidir.

$$s(n) = -\sum_{k=1}^p a_k s(n-k) + e(n) \quad (2.1)$$

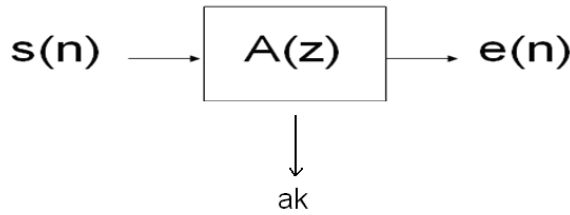
Konuşmayı; durağan kabul edildiği 20-30 ms'lik kısa sürelerde, oldukça başarılı bir şekilde kodlayabilen bu yöntemde,  $a_k$ ' lar; LPC katsayıları olarak adlandırılır ve bu süre zarfında (çerçeve süresi) sabittir. Konuşmanın gerçek değeri ile öngörülen değeri arasındaki fark; hata sinylidir:  $e(n)$  ve sentezlemede *uyarım* sinyali olarak adlandırılmaktadır (eşitlik (2.2)).

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (2.2)$$

$z$ -dönüşümü ile aktarım işlevi (transfer fonksiyonu) hesaplanırsa eşitlik (2.3) elde edilir.

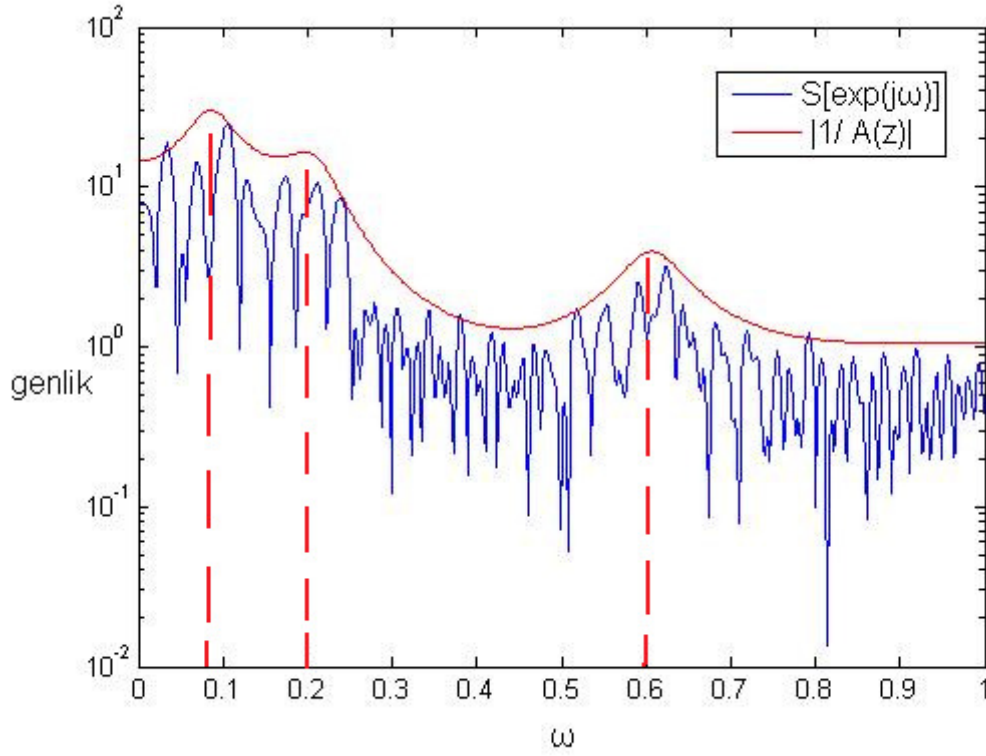
$$\frac{E(z)}{S(z)} = A(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (2.3)$$

Bu işlemler bize; sonlu uzunlukta dürtü tepkili [FIR, Finite Impulse Response] bir süzgeç (analiz filtresi) vermektedir. Böylece konuşma sinyali, incelendiği zaman aralığında, uyarım sinyaline ve süzgeç katsayılarına ( $a_k$ ) ayrıştırılmaktadır (şekil 2.2).



Şekil 2.2 LPC analiz süzgeci

Yukarıda bahsettiğimiz bu analiz süzgecinin tersi; sentez filtresi  $1/A(z)$  olarak bilinir. Sentez süzgecinin genlik tepkisinde yer alan tepeler; ses yolunun rezonans frekanslarına karşılık gelen *formant* frekansları olarak adlandırılmaktadır. Şekil 2.3'te formant frekanslarının bulunuşu anlatılmaktadır. İşte formant sentezleyiciler; konuşmayı bu ana bileşenleri (formant frekanslarını ve bant genişliklerini) ayarlayarak sentezlemeye çalışır.



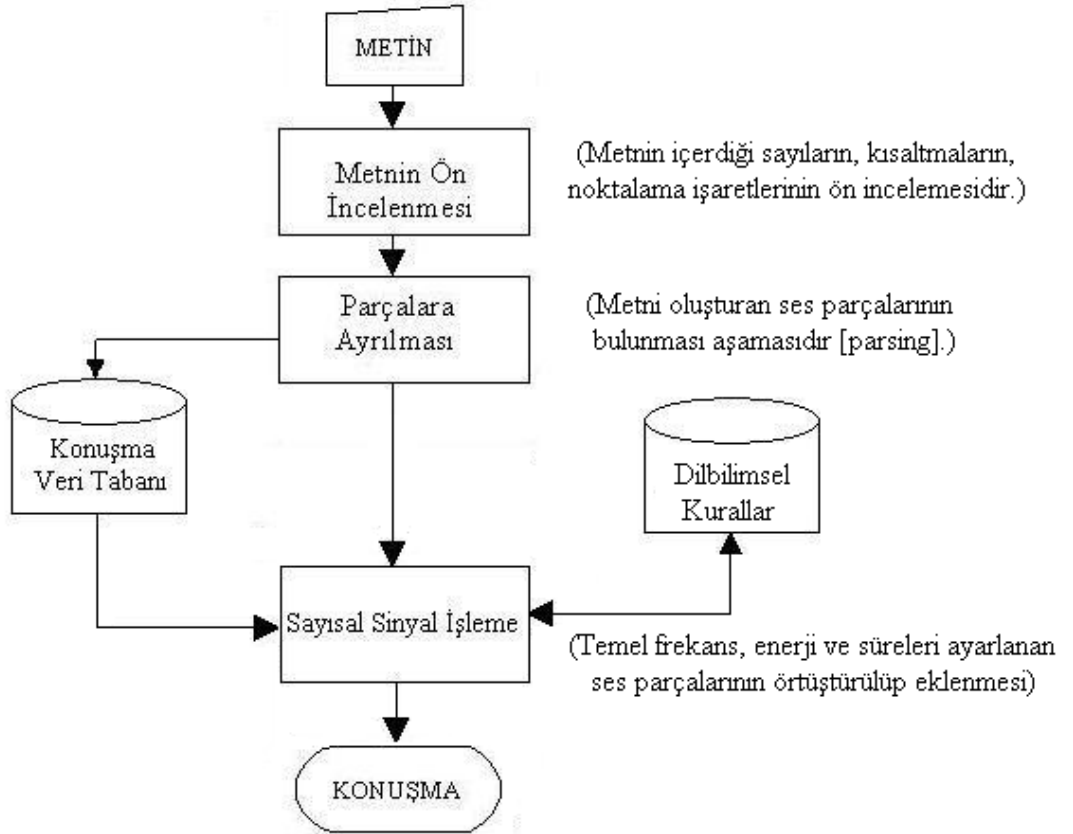
Şekil 2.3 Formant frekanslarının LPC sentez filtresinin genlik spektrumundan elde edilmesi

İkinci grupta yer alan *söyleyiş* sentezleyicilerinde amaç; tüm sesbirimlerin [phoneme], insanın ses üretim mekanizmasında oluşturulma şeklinin en doğru şekilde modellenmesidir. Tahmin edileceği üzere çok sayıda parametre içeren bu modelin elde edebildiği konuşma kalitesi formant sentezleyicilerinkinden daha iyidir. Fakat işlem ödünleşimi [tradeoff] de bir o kadar fazladır.

Üzerinde duracağımız 3. grup sentezleyiciler ise eklemeli [concatenative] sentezleyiciler olarak bilinir. Bu gruptaki sentezleyiciler; bellek kapasitelerinin ve sayısal sinyal işlemci hızlarının artışına paralel olarak tercih edilmektedir. Bunlar, konuşmayı; sesbirim, difon, trifen, seslem (hece) vb. gibi önceden kaydedilen ses parçalarını belirli sinyal işleme teknikleriyle biraraya getirerek oluşturmaya çalışırlar. Bu amaçla, ilk önce birleştirilecek parçaların veritabanı hazırlanmalıdır. Kayıt aşaması zahmetlidir. Hedef ses parçalarının kaydedilen konuşma parçalarından ayrılması işlemi de oldukça zaman alıcıdır. Bu işlemi belirli bir başarıyla otomatik olarak yapabilen

programlar vardır. Eklemeli sentezlemede yer alan temel adımlar şekil 2.4 ile gösterilmektedir.

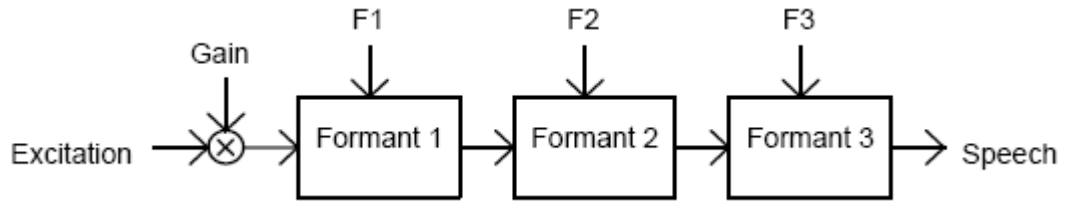
Her üç yöntem için de, ortak olan bölüm; kullanılacak olan ses parçalarının belirlenmesidir. Yazının dilbilgisel [grammar], biçimbirimsel [morphosyntactic] ve anlambilimsel [semantic] olarak incelenmesi aşamaları; DDİ alanının konusudur. Söz gruplarının belirlenmesi, sesteş sözcükler için belirsizliklerin çözülmesi, vurgu bilgilerinin çıkartılması gibi işlemler bu bölümde gerçekleşir.



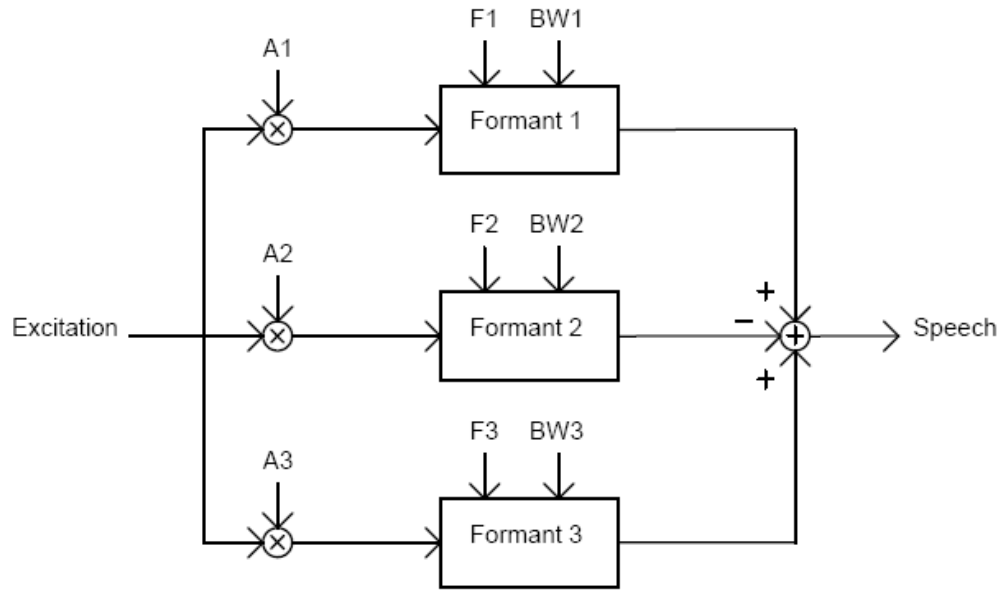
Şekil 2.4 Eklemeli sentezlemede işlemlerin akış şeması

### 2.2.1 Kural-tabanlı sentezleyiciler

Kural-tabanlı sentezleyicilerin çalışma prensibi konuşma organlarının modellenmesine (sistem modelleme) dayanır. Oldukça karmaşık yapıdadırlar. Değişik seslerin üretilme kuralları, uzun uğraşlar sonucu bulunmaya çalışılmıştır. Formant sentezleyiciler, bu grupta yer almaktadır. Seri ve paralel olmak üzere iki tür formant sentezleyici vardır ve şekil 2.5 ile verilmiştir.



a. Seri formant sentezleyici



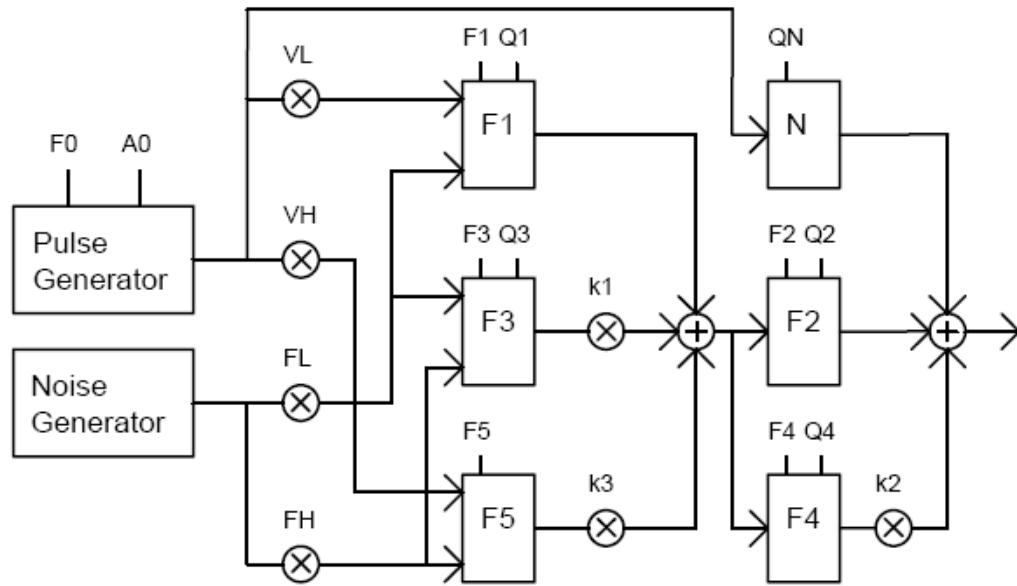
b. Paralel formant sentezleyici

Şekil 2.5.a. Seri formant sentezleyici, b. Paralel formant sentezleyici

Seri formant sentezleyiciler, denetim bilgisi olarak sadece formant frekanslarına ihtiyaç duyar. Ünlülerin sentezinde, formant frekanslarındaki genlik değerlerini ayrı ayrı

kontrol etmek gerekmez. Seri formant sentezleyicilerin sürtünücü ve patlamalı ünsüzleri sentezlemede çok başarılı olmadığı belirtilmektedir (Lemmetty 1999). Paralel yapıya göre daha az denetim bilgisine sahip olduğu için, daha kolay gerçekleştirilebilir.

Paralel formant sentezleyiciler, her bir formant frekansı için ayrı frekans, genlik ve bandgenişliği parametrelerine ihtiyaç duyar. Dolayısıyla daha karmaşık yapıdadır. Geniz (nasal) ünsüzlerinin ve sürtünücü ünsüzlerin sentezinde daha iyi sonuçlar verdiği, ancak ünlülerin sentezini gerçekleştiremediği belirtilmektedir (Lemmetty 1999). Bir diğer yaklaşım da seri ve paralel kombinasyonu kullanmaktır. Bu tip bir sentezleyici Şekil 2.6'da görülmektedir. PARCAS (Parallel-Cascade) modeli olarak adlandırılan bu formant sentezleyici Fince için tasarlanmıştır.



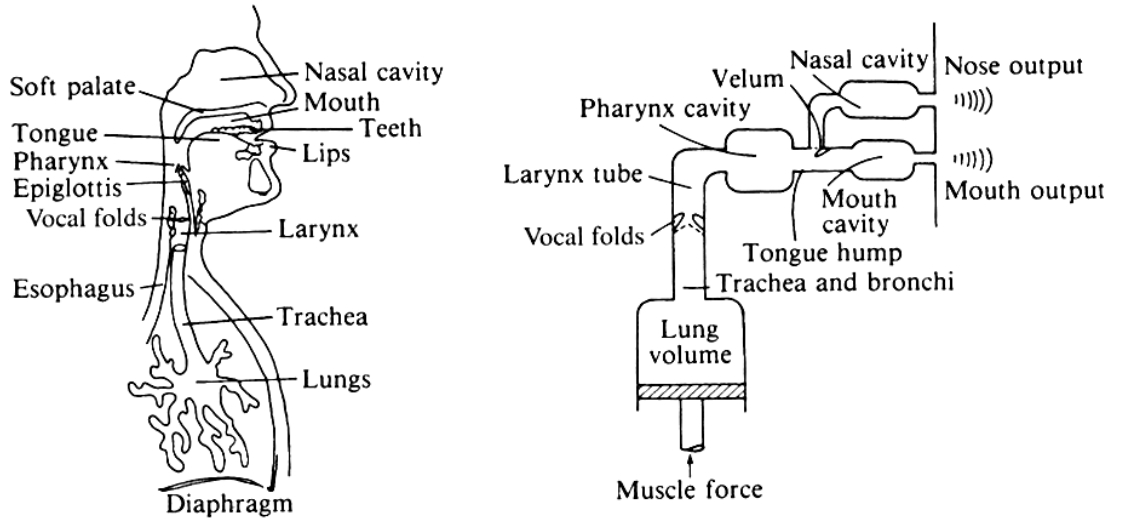
Şekil 2.6 PARCAS model seri-paralel formant sentezleyici

Bu grup sentezleyiciler, daha çok Dilbilimciler tarafından kullanılmaktadır. Çünkü bu sistemler, konuşmanın üretilmesiyle ve bilişsel yaklaşımlarla ilgili deneme imkânları sunar. En çok bilinen formant sentezleyiciler: Klatt Sentezleyici, DECtalk, MITalk (Allen vd. 1987), JSRU Sentezleyici ve INFOVOX'tur (Dutoit 1999). Formant

sentezleyiciler sonsuz sayıda ses üretebilme esnekliği gösterdiğinden, bu noktada eklemeli sentezleyicilerden avantajlı kabul edilirler (Pasanen 2001).

### 2.2.2 Söyleyiş sentezleyicileri

Söyleyiş sentezleyicileri; konuşma organlarımızı modellemeye çalışırlar. Ses tellerinde oluşan titreşimler, ağız, burun, çene ve dudaktan geçerken şekillendirilir. Söyleyiş sentezleyicileri bu organlarda gerçekleşen süreçleri modelleyerek ses sentezi yapmaya çalışır (şekil 2.7). Söyleyiş sentezleyicilerinin yapısı karmaşıktır, ancak doğal konuşma sentezinde başarılı sonuçlar vermektedir.



Şekil 2.7 Konuşma üretim mekanizması ve modellenmesi (Palo 2006)

### 2.2.3 Eklemeli sentezleyiciler

Bu grupta yer alan sentezleyiciler, konuşma sinyalini modellemeye çalışır. Eklemeli sentezleyiciler; önceden kaydedilmiş konuşma parçalarını birleştirerek sentezleme yaparlar. Birleştirilen konuşma parçaları: fonem, difon, trifon, seslem (hece) veya sözcük olabilir. Eklemeli sentezlemede genellikle difonlar tercih edilmektedir. Bunun sebepleri; hem uygun sayıda, örneğin Türkçe için yaklaşık 1600 adet, difon bulunması ve kayıt yükünün fazla olmaması, hem de difonların fonemler arası ses geçişlerini içermesidir (Karlı 2005). Bu çalışmada kullanılan difon veritabanının özellikleri, Materyal ve Yöntem adlı 3. Bölümde ayrıntılı olarak anlatılmıştır. Tez kapsamında ayrıca, yazılı metinlerdeki digram dağılımlarından difonların kullanım sıklıklarına geçiş yapılmış ve Türkçede en sık rastlanan difonların hangileri olduğu belirlenmiştir. Bu bilgiler de Ek 1 bölümünde ayrıntılı bir şekilde anlatılmaktadır.

### 2.3 Türkçe MKS Sistemleri İçin Literatür Tarama

Bu konuda ilk yapılan çalışmaların LPC tabanlı oldukları görülmektedir (Gerçek 1991, Büyükaşıkoglu 1991, Başara 1992). Daha sonra ve son yıllarda yapılan çalışmalarda eklemeli sentez yönteminin tercih edildiği görülmektedir (Özüm 1993, Ayhan 1998, Salor 1999, Bozkurt 2000, Özen 2002, Eker 2002, Görmez 2009, Erdemir 2010 ve Arık 2011). Hem sözcük hem de cümle için temel frekans konturunun değişimi ezgi modelleme çalışmalarının esas ilgi odağı olmuştur (Oskay 2002, Vural 2003 ve Öztürk 2005). Bunun yanı sıra Şayli (2002) ve Öztürk (2005), ses parçalarının sürelerini modellemeye çalışmışlardır. Derlem (corpus) tabanlı tek çalışma ise Sak (2004)'ın çalışmasıdır. Önür (2001), sinüzoidal modelleme ile konuşma sentezi üzerinde çalışmışken, Karlı (2005), prototip dalgaşekli ara değerlendirme yöntemiyle sentez yapmıştır. Aktan (2004), LPC kodlanmış harfleri uç uca birleştiren bir yonga tasarlamış, Ünalı (2007) ise, taşınabilir mobil cihazlar için MKS sistemi geliştirmiştir. Bunların yanı sıra Saklı Markov Modeli (SMM) tabanlı MKS sistemi çalışmaları bulunmaktadır (Güner ve Demiroğlu 2011, Orhan ve Demiroğlu, 2011).

Türkçe MKS için yapılmış olan lisansüstü çalışmalar çizelge 2.1’de verilmiştir.

Çizelge 2.1 Türkçe MKS için yapılmış olan lisansüstü çalışmalar

| No | Yazar               | Çalışmanın Türü, Yılı | Çalışmanın Başlığı                                                                                           | Yöntem / Özellik                                                                       |
|----|---------------------|-----------------------|--------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| 1  | Alper GERÇEK        | Yüksek Lisans, 1991   | “A TMS 5220 based speech synthesis development system”                                                       | Doğrusal öngörümlü kodlama ve donanım tabanlı                                          |
| 2  | Karen BÜYÜKAŞIKOĞLU | Yüksek Lisans, 1992   | “Konuşma işaretlerinin analiz ve sentezi”                                                                    | Doğrusal öngörümlü kodlama kullanılmıştır                                              |
| 3  | Enis Sezai BAŞARA   | Yüksek Lisans, 1992   | “Yapay ses üretim yöntemleri”                                                                                | Doğrusal öngörümlü kodlama kullanılmıştır                                              |
| 4  | İlhan Yaşar ÖZÜM    | Yüksek Lisans, 1993   | “A speech synthesis system for Turkish language based on the concatenation of phonemes taken from a speaker” | Fonem ve difon birleştirme (doğrusal aradeğerleme ile), TMS 320 C25 ile gerçekleştirme |
| 5  | Kamil GÜVEN         | Yüksek Lisans, 1994   | “PC based speech synthesis for Turkish”                                                                      | formant ses sentezleyici ile Türkçe ünlü harflerin sentezlenmesi                       |
| 6  | Murat Servet ERER   | Yüksek Lisans, 1994   | “Karma söz üretme yöntemi ile Türkçe yazılı metinden söze geçme”                                             | Hece tabanlı iki farklı birleştirme yöntemi                                            |
| 7  | Kerem AYHAN         | Yüksek Lisans, 1998   | “Text to speech synthesizer in Turkish using non parametric techniques”                                      | Hece birleştirme (ünlü birleşme noktalarında), PSOLA ve türevleri ile                  |
| 8  | Özgül SALOR         | Yüksek Lisans, 1999   | “Signal processing aspects of text to speech synthesizer in Turkish”                                         | Hece (ünsüz birleşme noktalarında), LP-PSOLA (REL) yöntemiyle                          |
| 9  | Barış BOZKURT       | Yüksek Lisans, 2000   | “Reading aid for visually impaired (A Turkish text-to-speech system development)”                            | TD-PSOLA                                                                               |
| 10 | Ömer ESKİDERE       | Yüksek Lisans, 2000   | “Yazılım tabanlı söz sentezleyici tasarımı”                                                                  | formant sentezleme tekniğine dayanan bir Türkçe sözcük sentezleyici                    |
| 11 | Çağla ÖNÜR          | Yüksek Lisans, 2001   | “Concatenative speech synthesis based on a sinusoidal speech model”                                          | Sinüzoidal model                                                                       |
| 12 | Erkan ABDULLAHBEŞE  | Yüksek Lisans, 2001   | “Fundamental frequency contour synthesis for Turkish text to speech”                                         | Sözdizimsel özelliklere dayalı $F_0$ eğri benzetimi                                    |
| 13 | Şifa Serdar ÖZEN    | Yüksek Lisans, 2002   | “Türkçe metinden konuşma sentezleme”                                                                         | LP-PSOLA yöntemiyle difon birleştirme                                                  |
| 14 | Barış EKER          | Yüksek Lisans, 2002   | “Turkish text to speech system”                                                                              | Ötümlü bölümlerde difon birleştirme, TD-PSOLA yöntemiyle                               |

Çizelge 2.1 Türkçe MKS için yapılmış olan lisansüstü çalışmalar (devam)

|    |                |                     |                                                                                                                                                            |                                                                                                                   |
|----|----------------|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| 15 | Ömer ŞAYLI     | Yüksek Lisans, 2002 | “Duration analysis and modeling for Turkish text-to-speech synthesis”                                                                                      | Doğrusal-toplamsal süre modeli, ortalama fonem ve trifon süreleri, trifon ağaç modeli                             |
| 16 | Banu OSKAY     | Yüksek Lisans, 2002 | “Automatic modelling of Turkish prosody”                                                                                                                   | Tümce vurgusu                                                                                                     |
| 17 | Esra VURAL     | Yüksek Lisans, 2003 | “A prosodic Turkish text-to-speech synthesizer”                                                                                                            | Festival sistemiyle, sözcük vurgusu ve allofon süreleri                                                           |
| 18 | Haşim SAK      | Yüksek Lisans, 2004 | “A corpus based concatenative speech synthesis system for Turkish”                                                                                         | Derlem tabanlı, harmonik kodlama                                                                                  |
| 19 | Ozan AKTAN     | Yüksek Lisans, 2004 | “A single chip solution for text-to-speech synthesis”                                                                                                      | LPC kodlanmış harf kayıtlarının tek bir yonga ile uç uca birleştirilmesi                                          |
| 20 | Asude KARLI    | Yüksek Lisans, 2005 | “Örnek bir dizi cümle için Türkçe metinden konuşma sentezleyici”                                                                                           | Prototip dalgaşekli aradeğerleme yöntemiyle difon birleştirme                                                     |
| 21 | Özlem ÖZTÜRK   | Doktora, 2005       | “Modeling phoneme durations and fundamental frequency contours in Turkish speech”                                                                          | Bağlanım ağaçları (regression trees)                                                                              |
| 22 | İlker ÜNALDI   | Yüksek Lisans, 2007 | “Taşınabilir cihazlar için Türkçe metinden konuşma sentezleme sistemi”                                                                                     | LP-PSOLA yöntemiyle difon birleştirme (mobil cihaz üzerinde)                                                      |
| 23 | Cansel DEMİR   | Yüksek Lisans, 2009 | “Konuşma tanıma sentezleme sistemlerinin okul öncesi dönem yabancı dil eğitiminde kullanılması”                                                            | Microsoft Speech Server TTS kullanılmıştır                                                                        |
| 24 | Kenan GÜLDALI  | Yüksek Lisans, 2009 | “Türkçe metin seslendirme”                                                                                                                                 | Hece birleştirme kuralları belirlenmiştir                                                                         |
| 25 | Zeliha GÖRMEZ  | Yüksek Lisans, 2009 | “Implementation of a text-to-speech system with machine learning algorithms in Turkish”                                                                    | Birim seçme tabanlı TD-PSOLA ile eklemeli sentez                                                                  |
| 26 | Yücel BİCİL    | Yüksek Lisans, 2010 | “Türkçe metinden konuşma sentezleme”                                                                                                                       | Difon ve trifonlar kullanarak TD-PSOLA ile eklemeli sentez                                                        |
| 27 | Cavit ERDEMİR  | Yüksek Lisans, 2010 | “Türkçe metin seslendirme için doğal konuşma sentezleme”                                                                                                   | Dalgışekli birleştirme yöntemiyle ve “en çok iki harfli hece” parçacık birimi temel alınarak sentez yapılmaktadır |
| 28 | Tuncay ŞENTÜRK | Yüksek Lisans, 2010 | “Türkçe metin seslendirme”                                                                                                                                 | çift-ses ekleme, hece ekleme, ses süre ve şiddetinin değiştirilmesi                                               |
| 29 | Güray ARIK     | Yüksek Lisans, 2011 | “Görme engelliler için bilgisayar kullanımının etkinleştirilmesi, erişilebilirlik ve bir Türkçe hece tabanlı konuşma sentezleme sisteminin geliştirilmesi” | Hece tabanlı bir eklemeli konuşma sentezleyici tasarlanmıştır                                                     |

Takip eden bölümde, tez çalışmasında kullanılan materyalden ve uygulanan yöntemlerden bahsedilecektir. 4. Bölümde incelenecek olan ezgi modelimizin dayandığı akustik özellikler ve bunlar üzerindeki sinyal işleme yöntemleri takip eden 3. bölümde anlatılmıştır.

### 3. MATERYAL VE YÖNTEM

Bu bölümde, tezde kullanılan difon veritabanının özellikleri, difonları bir araya getirmek ve difonlar üzerinde istenen ezgisel değişiklikleri yapmak için kullanılan sinyal işleme yöntemleri anlatılmaktadır.

#### 3.1 Difon Veritabanı

Tezde birleştirilecek ses parçaları için ikili sesbirimler (difonlar) kullanılmıştır. Difonlar, Dr. Özgül Salor'un çalışmasından (Salor vd. 2002) alınmıştır. Türkçe için difon listesi çizelge 3.1'de verilmiştir.

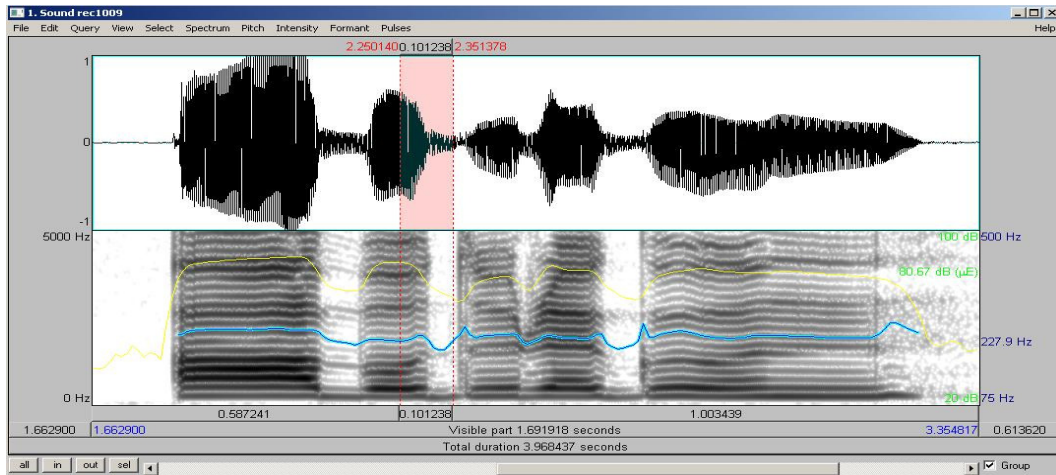
Çizelge 3.1 Difon listesi

| No       | IPA      | METUbet | Benim Fonem Listem | Örnek              | Açıklama    |
|----------|----------|---------|--------------------|--------------------|-------------|
| ÜNLÜLER  |          |         |                    |                    |             |
| 1        | A        | AA      | a1                 | <i>a1</i> nı       | artdamaksıl |
| 2        | A        | A       | a2                 | <i>l a2</i> f      | öndamaksıl  |
| 3        | E        | EE      | e1                 | <i>d e2 r e1</i>   | açık        |
| 4        | E        | E       | e2                 | <i>e2 l</i> ma     | kapalı      |
| 5        | İ        | I       | ı                  | <i>ı s ı</i>       | -           |
| 6        | İ        | IY      | i1                 | <i>i1</i> ğde      | açık        |
| 7        | <u>I</u> | IY      | i2                 | <i>s i2 m i2</i> t | kapalı      |
| 8        | <u>u</u> | O       | o1                 | <i>s o1</i> ru     | açık        |
| 9        | O        | O       | o2                 | <i>o2</i> ğlak     | kapalı      |
| 10       | <u>œ</u> | OE      | ö1                 | <i>ö1</i> rtü      | açık        |
| 11       | <u>ø</u> | OE      | ö2                 | <i>ö2</i> ğren     | kapalı      |
| 12       | u        | U       | u1                 | <i>k u1</i> lak    | açık        |
| 13       | U        | U       | u2                 | <i>u2</i> ğur      | kapalı      |
| 14       | y        | UE      | ü1                 | <i>ü1</i> mit      | açık        |
| 15       | Y        | UE      | ü2                 | <i>d ü2</i> ğme    | kapalı      |
| ÜNSÜZLER |          |         |                    |                    |             |
| 16       | b        | B       | b                  | <i>b</i> al        | -           |
| 17       | ɗ        | C       | c                  | <i>c</i> am        | -           |
| 18       | tʃ       | CH      | ç                  | <i>se ç im</i>     | -           |
| 19       | D        | D       | d                  | <i>d e d e</i>     | -           |
| 20       | f        | F       | f                  | <i>f</i> asıl      | -           |
| 21       | ʃ        | G       | g1                 | <i>g1</i> enç      | ön          |
| 22       | g        | GG      | g2                 | kar <i>g2</i> a    |             |
| 23       | :        | GH      | ğ                  | dü <i>ğ</i> me     |             |

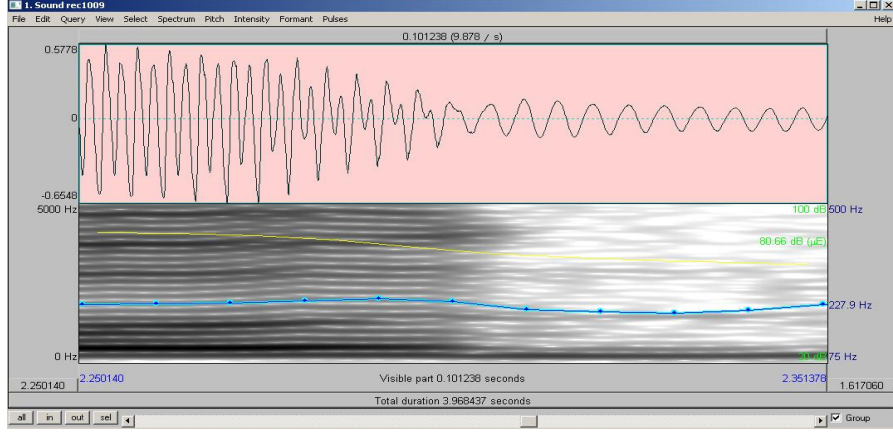
Çizelge 3.1 Difon listesi (devam)

|    |   |    |    |                         |                 |
|----|---|----|----|-------------------------|-----------------|
| 24 | h | H  | H  | <i>h</i> asta           | -               |
| 25 | ɟ | J  | j  | mü <i>j</i> de          | -               |
| 26 | c | K  | k1 | <i>k1</i> edi           | ön              |
| 27 | k | KK | k2 | a <i>k2</i> ıl          |                 |
| 28 | l | L  | l1 | <i>l1</i> a <i>l1</i> e |                 |
| 29 | ɭ | LL | l2 | ku <i>l2</i>            |                 |
| 30 | m | M  | m  | da <i>m</i>             | -               |
| 31 | n | NN | n1 | a <i>n1</i> ı           | dilucu-dişeti   |
| 32 | ɳ | N  | n2 | sü <i>n2</i> gü         | dil-artdamak    |
| 33 | p | P  | p  | i <i>p</i>              | -               |
| 34 | r | R  | r1 | <i>r1</i> af            | çok vuruşlu     |
| 35 | ɽ | RR | r2 | ı <i>r2</i> mak         | tek vuruşlu     |
| 36 | ɣ | RH | r3 | bi <i>r3</i>            | ötümsüz, sızıcı |
| 37 | s | S  | s  | <i>s</i> e <i>s</i>     | -               |
| 38 | ʃ | SH | ş  | a <i>ş</i> ı            | -               |
| 39 | t | T  | t  | ü <i>t</i> ü            | -               |
| 40 | v | VV | v1 | <i>v1</i> ar            | dudak-diş       |
| 41 | ʋ | V  | v2 | ta <i>v2</i> uk         | yarı ünlü       |
| 42 | j | Y  | y1 | <i>y1</i> at            |                 |
| 43 | ɰ | Y  | y2 | hu <i>y2</i>            |                 |
| 44 | z | Z  | z1 | a <i>z1</i> ık          | ötümlü          |
| 45 | ʒ | ZH | z2 | yo <i>z2</i>            | ötümsüz         |

16 kHz ile örneklenmiş ve “.wav” formatında saklanmış olan difon kayıtları, aşağıda şekil 3.1’de anlatıldığı gibi taşıyıcı ifadeden seçilerek şekil 3.2’deki gibi ayrıştırılmıştır. Difon veritabanında toplam 2284 adet difon bulunmaktadır.



Şekil 3.1 Taşıyıcı cümle içerisinde kaydedilmiş olan “o d ” difonu

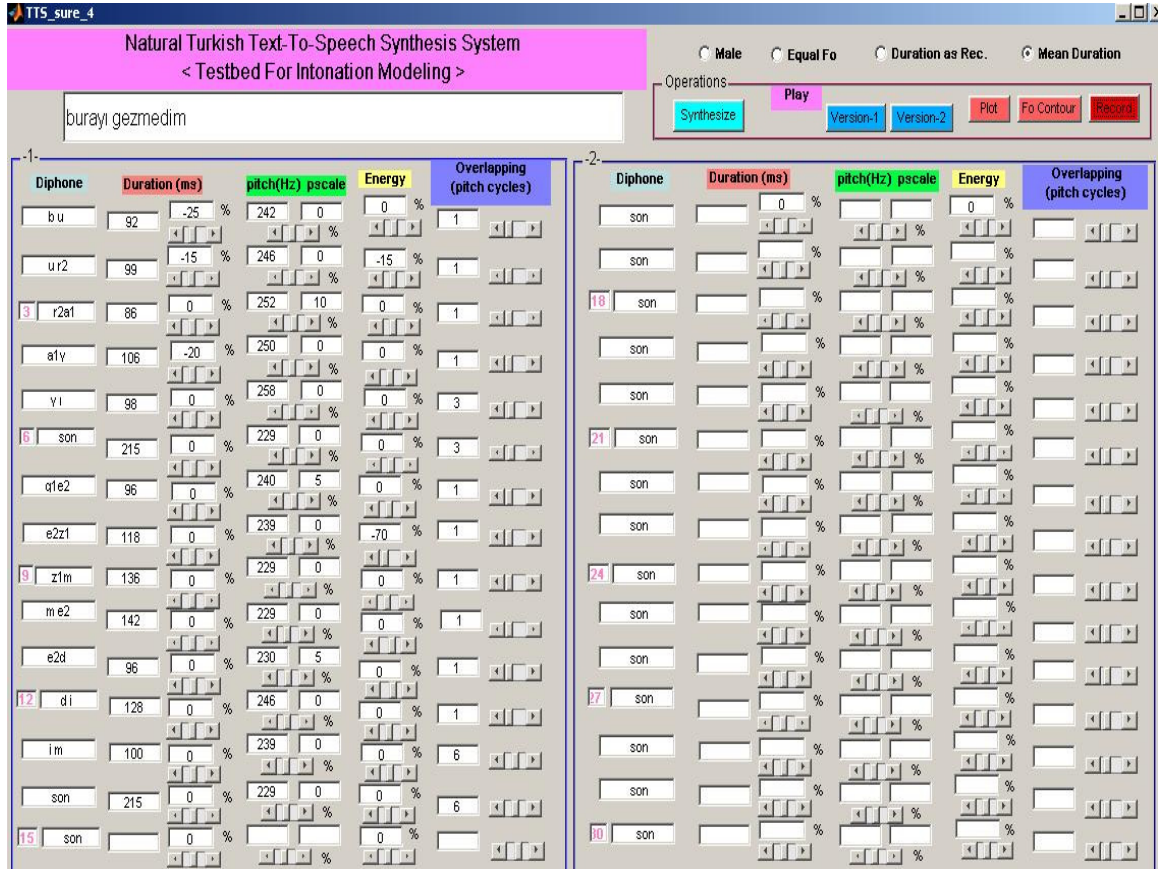


Şekil 3.2 Elden geçirilmiş ve ayrıştırılmış olan “o d” difonu dalgaşekli

Konuşma sentezlemede kullanılan süre ve perde ayarlarında, ötümlü-ötümsüz ayrımı ve perde periyodu tahmini gerekmektedir. Henüz otomatik yöntemi bulunamamış olan bu problemlerin çözümü için; perde işaretleri [pitch marks] elle belirlenmekte ve perde periyodu da bu işaretlerden yararlanılarak (ardışık perde işaretleri arasındaki fark kullanılarak) hesaplanmaktadır.

Ötümlü/ ötümsüz ayrımı için; bir perdelik çerçevelerin enerji, sıfır-geçiş sayısı, bir örnek gecikmeli özilinti katsayısı ve izgesel düzgünlük oranı [spectral flatness measure] kullanılmaktadır.

PSOLA yöntemiyle difonların birleştirilmesi ve difonlar üzerinde gerekli süre, perde frekansı ve enerji değişikliklerinin yapılabilmesi için aşağıda şekil 3.3'te gösterilen kullanıcı ara yüzü geliştirilmiştir.



Şekil 3.3 Matlab GUI’de geliştirilen MKS arayüzü

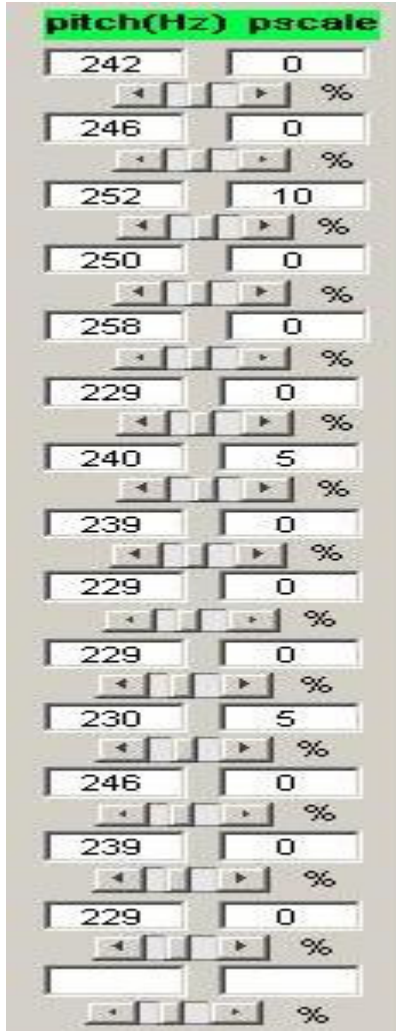
Geliştirilen arayüzdeki özellikler: difon belirleme, süre ayarı, perde frekansı ayarı, enerji ayarı ve örtüşürme perde sayısı (süresi) şeklindedir. Bu özelliklerin büyütülmüş halleri aşağıda şekil 3.4 – 3.6 ile verilmiştir.

-1-

| Diphone | Duration (ms) |       |
|---------|---------------|-------|
| b u     | 92            | -25 % |
| u r2    | 99            | -15 % |
| 3 r2a1  | 86            | 0 %   |
| a1v     | 106           | -20 % |
| y l     | 98            | 0 %   |
| 6 son   | 215           | 0 %   |
| q1e2    | 96            | 0 %   |
| e2z1    | 118           | 0 %   |
| 9 z1m   | 136           | 0 %   |
| m e2    | 142           | 0 %   |
| e2d     | 96            | 0 %   |
| 12 d i  | 128           | 0 %   |
| i m     | 100           | 0 %   |
| son     | 215           | 0 %   |

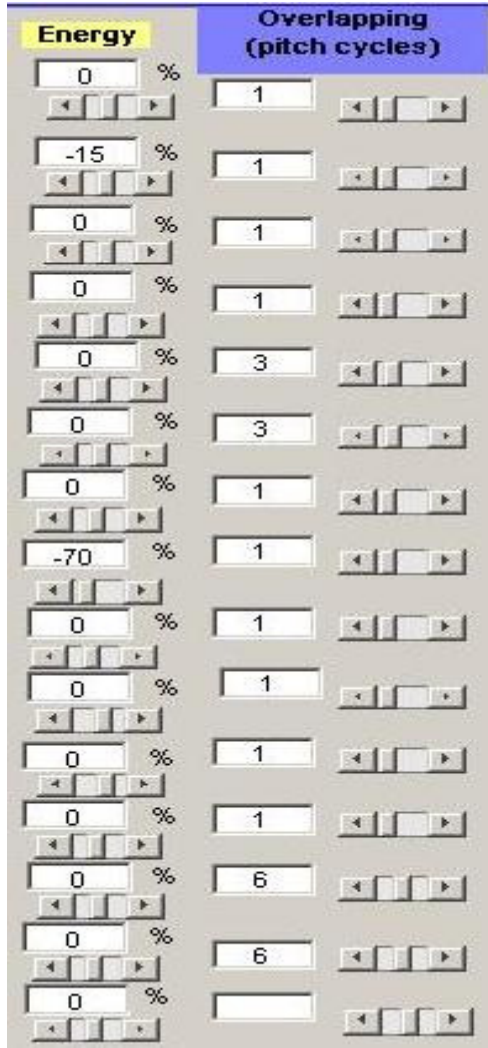
Şekil 3.4 Geliştirilen arayüzdeki süre ayarı özelliği

Şekil 3.4'te de görüldüğü üzere, süre yüzde olarak artırılıp azaltılabilir. PSOLA ile yapılan süre değiştirme işleminde, yeni perde işaretleri Bölüm 3.3.1'de anlatıldığı gibi belirlenmektedir.



Şekil 3.5 Geliştirilen arayüzdeki perde frekansı ayarı özelliği

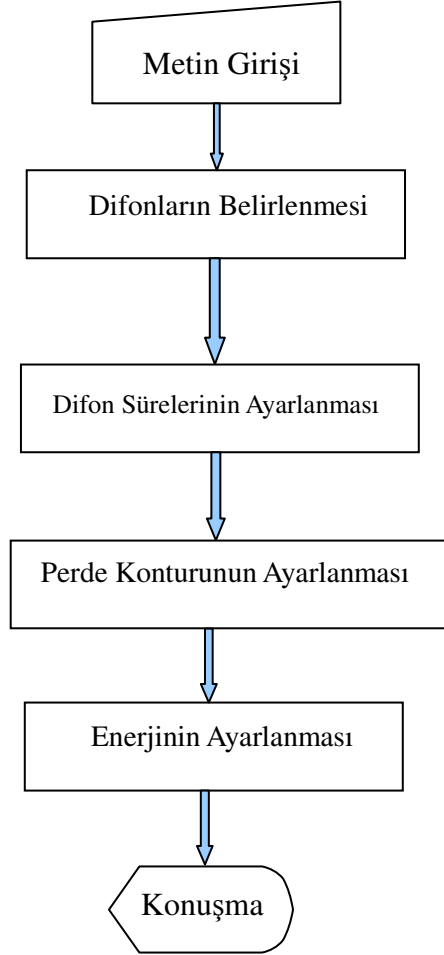
Şekil 3.5'te de görüldüğü üzere, perde frekansı yüzde olarak artırılıp azaltılmakta, bu işlem PSOLA yöntemiyle yapılmaktadır (bkz. Bölüm 3.2.3).



Şekil 3.6 Geliştirilen arayüzdeki enerji ve örtüşme perde sayısı ayarı özelliği

Şekil 3.6’da da görüldüğü üzere, enerji yüzde olarak değiştirilmekte, ayrıca örtüşme perde sayısı ayarlanabilmektedir. Tez çalışmasında, yapılan pek çok denemenin yardımıyla ötümlü ses birleşimlerinde 3 perde, ötümsüz ses birleşimlerinde 1 perde ve boşluk ile birleştirirken 6 perde örtüştürme yapılması tercih edilmiştir. Burada önemli olan nokta; ötümsüz ses birleşmelerinde fazla perde örtüştürme yapıldığında, ses geçişlerinin birbiri içine girerek duyulmaması durumudur. Bu şekilde başarılı bir sentez elde edilmemektedir. Örnek olarak: “tepsi” sözcüğünü sentezlerken, “ep” ve “ps” difonları birleştirilirken, “s” sesinde bir kayıp olmaması için yalnızca 1 perde örtüştürme yapılmalıdır. Ayrıntılar 3.3 bölümünde ele alınmaktadır.

Şekil 3.3'te verilen arayüzün arkaplânında çalışan programların akış şeması şekil 3.7 ile verilmiştir.



Şekil 3.7 Geliştirilen arayüzde çalışan işlemlerin akış şeması

Girilen metin önce sözcüklere, daha sonra hecelere ve difonlara ayrılmaktadır. Doğru difonların belirlenmesi için, aşağıda çizelge 3.2 ile verilen ve Ergenç (2002) referans alınarak tanımlanan kurallar dikkate alınmaktadır. Daha sonra difonların süreleri ayarlanır. Bu yapılırken, sözcük ve cümle uzunluğuna göre belirlenen süre kısaltma işlemi uygulanır (bu işlemin ayrıntıları Bölüm 4.4'te anlatılmaktadır). Sentezin

özelliklerine göre, perde konturunun belirlenmesinin ardından, perde frekansı değiştirme ve son olarak enerji değiştirme işlemleri uygulanır ve konuşma sentezlenir.

Burada difonların belirlenmesi aşamasında, birden fazla türü olan fonemlerin doğru olanının bulunması için aşağıda çizelge 3.2’de verilen kurallar uygulanmıştır.

Çizelge 3.2 Sesbirim belirleme kural ve örnekleri

| Fonem | Çeşitleri | Sesbirim belirleme kuralı                                | Örnek                                                                   |
|-------|-----------|----------------------------------------------------------|-------------------------------------------------------------------------|
| g     | g1        | öndil ünlüleri ile birlikte ise “g1”                     | bel <b>g1</b> e, sevg <b>1</b> ili, g <b>1</b> ömüt, övg <b>1</b> ü     |
|       | g2        | arkadil ünlüleri ile birlikte ise “g2”                   | <b>g2</b> azete, yarg <b>2</b> ı, g <b>2</b> oril, sorg <b>2</b> ucu    |
| k     | k1        | <i>öndil ünlüleri</i> ile birlikte ise “k1”              | <b>k1</b> edi, tep <b>k1</b> i, k <b>1</b> ömür, türk <b>1</b> ücü      |
|       | k2        | <i>arkadil ünlüleri</i> ile birlikte ise “k2”            | hak <b>2</b> an, şark <b>2</b> ı, k <b>2</b> utu, sak <b>2</b> lamak    |
| l     | l1        | öndil ünlüleri ile birlikte ise “l1”                     | kel <b>1</b> ebek, siml <b>1</b> i, l <b>1</b> ökosit, bül <b>1</b> bül |
|       | l2        | arkadil ünlüleri ile birlikte ise “l2”                   | sakl <b>2</b> amak, yanl <b>2</b> ı, kol <b>2</b> onya, tuzl <b>2</b> u |
|       |           | başta ise “l1”                                           | <b>l1</b> oş, l <b>1</b> unapark                                        |
| n     | n1        | hemen hepsi “n1”                                         | an <b>1</b> ı, kon <b>1</b> ser                                         |
|       | n2        | n’den sonra g veya y geliyorsa “n2”                      | sün <b>2</b> gü, dün <b>2</b> ya                                        |
| r     | r1        | başta ise “r1”                                           | <b>r1</b> ozet, r <b>1</b> ay                                           |
|       | r2        | ortada ise “r2”                                          | tar <b>2</b> ak, sor <b>2</b> un                                        |
|       | r3        | sonda ise “r3”                                           | kar <b>3</b> , kibar <b>3</b>                                           |
| v     | v1        | başta veya ortada ise “v1”                               | sav <b>1</b> cı, v <b>1</b> ana                                         |
|       | v2        | sonda ise “v2”                                           | sav <b>2</b> , kov <b>2</b>                                             |
|       |           | v’den sonra “u” geliyorsa “v2”,<br>diğerleri için “v1”   | tav <b>2</b> uk, dav <b>2</b> ul<br>tav <b>1</b> a, sıv <b>1</b> acı    |
| y     | y1        | başta ise “y1”                                           | <b>y1</b> atak, y <b>1</b> etim                                         |
|       | y2        | sonda ise “y2”                                           | huy <b>2</b> , aday <b>2</b>                                            |
|       |           | y’den sonra ünlü geliyorsa “y1”,<br>ünsüz geliyorsa “y2” | gay <b>1</b> e, kuy <b>1</b> u<br>hay <b>2</b> van, tay <b>2</b> fun    |
| z     | z1        | başta veya ortada ise “z1”,                              | koz <b>1</b> a, z <b>1</b> amir                                         |
|       | z2        | sonda ise “z2”                                           | koz <b>2</b> , kiraz <b>2</b>                                           |

## 3.2 MKS Sistemleri İçin Sinyal İşleme Yöntemleri

Bu bölümde konuşma sentezinde kullanılan farklı sinyal işleme yaklaşımları ele alınmaktadır. Sinüsoidal modelleme, HNM, TD-PSOLA, LP-PSOLA ve MBROLA yöntemleri incelenmiştir. Tezde kullanılan yöntem TD-PSOLA olduğu için, bu yöntem ayrıntılı olarak Bölüm 3.3'te anlatılmıştır.

### 3.2.1 Sinüsoidal modelleme

Periyodik sinyalleri Fourier seri açılımı ile ifade edebiliriz. Sinüsoidal modelleme sonlu sayıda sinüsoidalle çerçeveleri ifade etmeye çalışır (Eşitlik 3.1).

$$\hat{s}(n) = \sum_{l=1}^L A_l \cos(l\omega_0 n + \phi_l) \quad (3.1)$$

Burada;

$A_l$ : sinüsoidallerin genlikleri

$\omega_0$ : temel açısal frekans

$\phi_l$ : sinüsoidallerin fazlarıdır.

### 3.2.2 HNM (Harmonic plus Noise Model)

Bu yöntemde konuşma sinyali harmonik ve gürültü olmak üzere iki bileşene ayrıştırılır. Harmonik bölümü konuşmanın neredeyse periyodik olan bölümüdür. Bu harmonikler temel frekansın katları şeklinde olup genlik ve fazları kestirilmeye çalışılır. Gürültü bileşeni ise sinyalden harmonik bileşenin farkının alınmasıyla bulunur.

Harmonik bileşen:

$$h(n) = \sum_{k=0}^L a_k(n) \cos(\phi_k(n)) \quad (3.2)$$

Gürültü bileşeni:

$$r(n) = s(n) - h(n) \quad (3.3)$$

Bu yöntemde prosodik değişiklikler yalnızca harmonik bölümde yapılır.

### **3.2.3 TD-PSOLA**

TD-PSOLA (Time Domain – Pitch Synchronous Overlap and Add) yöntemi bu tez çalışmasında kullanılan yöntemdir ve Bölüm 3.3’te ayrıntılı olarak incelenecektir.

### **3.2.4 LP-PSOLA**

LP-PSOLA (Linear Prediction – Pitch Synchronous Overlap and Add) yönteminde, perde işaretleri tarafından belirlenen çerçevelerin doğrusal öngörümlü kodlama ile kalıntı [residual] sinyalleri bulunur ve temel frekans değiştirme, süre değiştirme gibi işlemler kalıntı sinyalleri üzerinde yapılır.

### **3.2.5 MBROLA**

Bu yöntemde amaç; TD-PSOLA yönteminde karşılaşılan faz, temel frekans ve izge uyumsuzluğu gibi olumsuzlukların giderilmesidir. Bu amaçla, veritabanında bulunan tüm ses parçalarının ötümlü bölümleri, her perde için, aynı temel frekansa, izgel zarfları da rastgele fakat sabit ilk fazlara getirilir. Böylece örtüştürüp ekleme yapılırken, faz veya temel frekans uyumsuzluğu olmayacaktır. Bu yöntemde, perde işaretlerinin bulunması da gerekmemektedir (Dutoit ve Leich 1993). MBROLA yöntemi ile TD-PSOLA’ya göre daha kaliteli sentez yapıldığı rapor edilmiştir.

MBROLA Projesi TCTS laboratuvarında, Faculte Polytechnique de Mons, Belçika’da yapılmıştır. Projenin amacı; çok dilli konuşma sentezleme araştırmalarını yaygınlaştırmak, özellikle prosodi üretimi çalışmalarını artırmaktır (Dutoit vd. 1996).

### 3.3 TD-PSOLA Yöntemi

TD-PSOLA yönteminde, birleştirilecek olan parçalar perde işaretleri referans alınarak belirlenir. PSOLA yöntemi, CNET (the National Centre of Telecommunications Studies) tarafından geliştirilmiştir. Yöntemin matematiksel ifadesi eşitlik (3.4)–(3.6), çizimsel gösterimi ise şekil 3.8 ile verilmektedir. Bu yöntemde, birleştirilecek olan konuşma parçaları, perde işareti (pitch marker) adı verilen referans noktaları dikkate alınarak birleştirilir. Böylece birleştirilecek olan sinyallerin faz uyumu sorunu ortadan kalkmış olur.

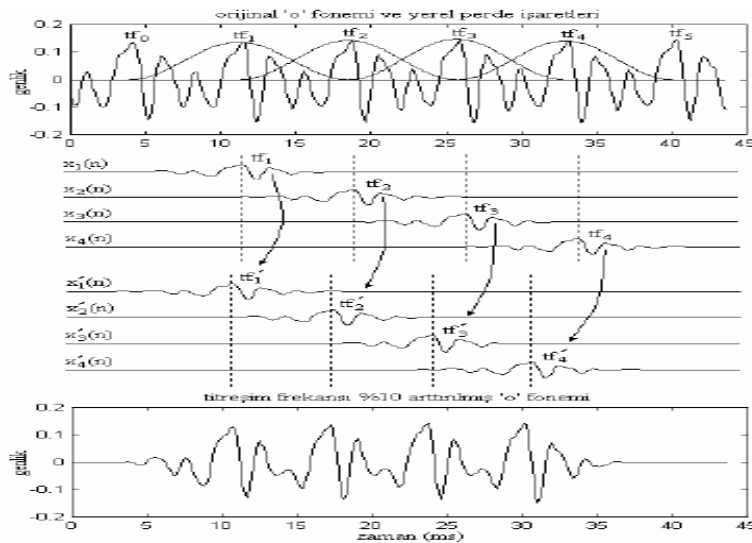
$$s_i(n) = s(n)w(n - iT_0) \quad (3.4)$$

$$\tilde{s}(n) = \sum_{i=-\infty}^{\infty} s_i(n - i(T - T_0)) \quad (3.5)$$

$$\tilde{s}(n) = \sum_{i=-\infty}^{\infty} s(n - i(T - T_0))w(n - iT) \quad (3.6)$$

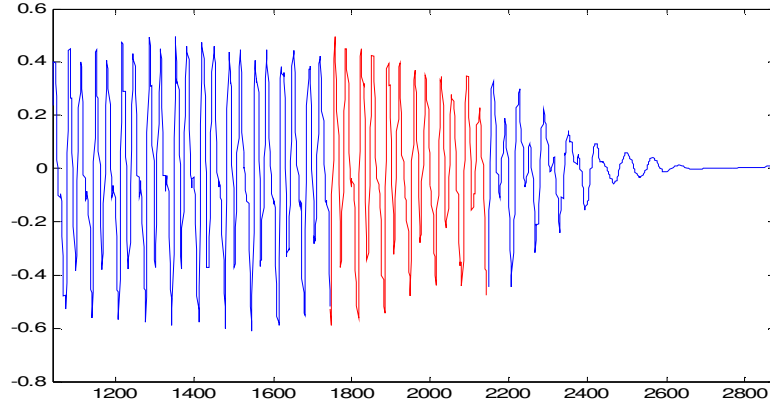
Burada  $T_0$ : orijinal temel periyot,  $T$  ise hedeflenen yeni temel periyottur.

$w(n)$  ise çerçeveleri ağırlıklandırmada kullanılan penceredir. Hamming, Hanning veya Bartlett pencere olabilir.

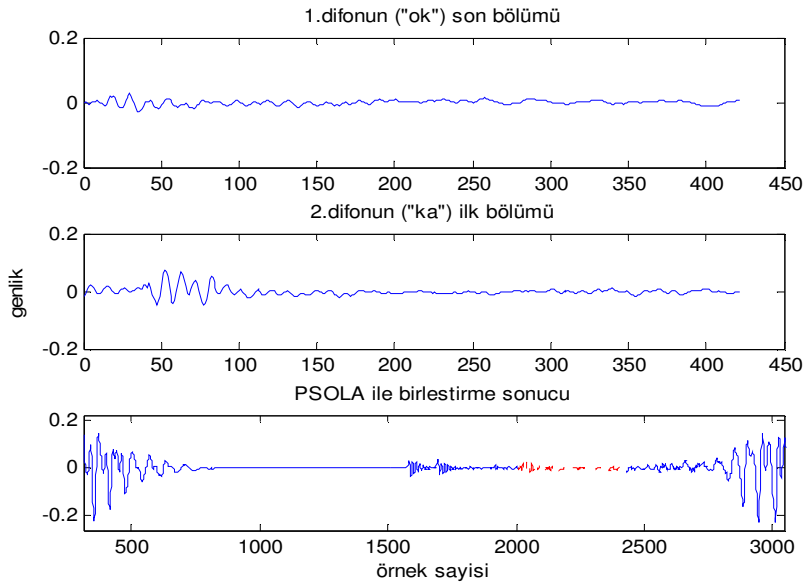


Şekil 3.8 PSOLA yöntemiyle örtüştürüp ekleme (Özen 2002)

Aşağıda şekil 3.9’da, PSOLA yöntemiyle “b a1” + “a1k2” birleşmesi (ünlü-ünlü), şekil 3.10’da “o k2” + “k2a1” birleşmesi (ünsüz-ünsüz) gösterilmiştir.



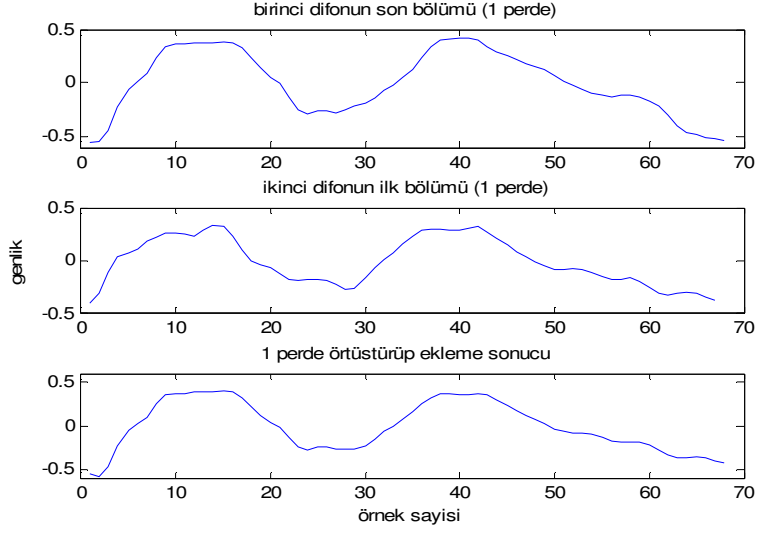
Şekil 3.9 “b a1” + “a1k2” difonlarının PSOLA ile birleştirilmesi



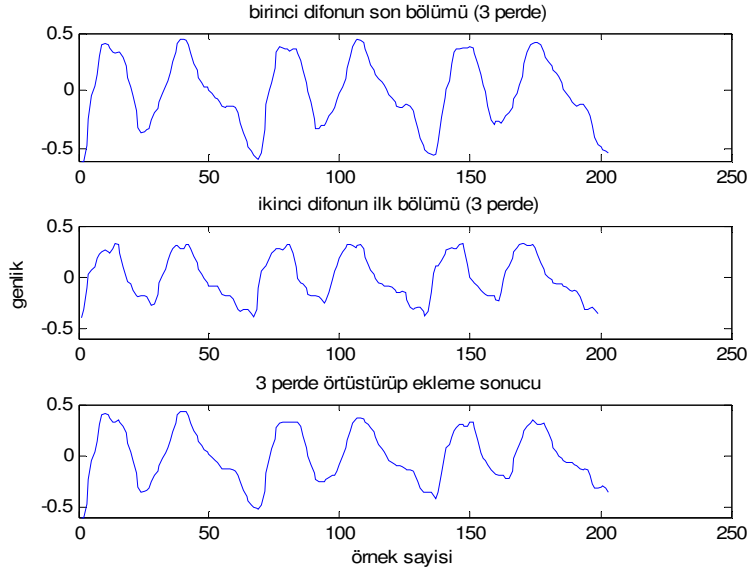
Şekil 3.10 PSOLA yöntemiyle “o k2” + “k2a1” birleşmesi

Tez çalışmasında, PSOLA ile eklemede örtüştürme süresi değişken alınarak sentez kalitesine olan etkisi incelenmiştir. GUI’de örtüştürme miktarı bağımsız bir parametre

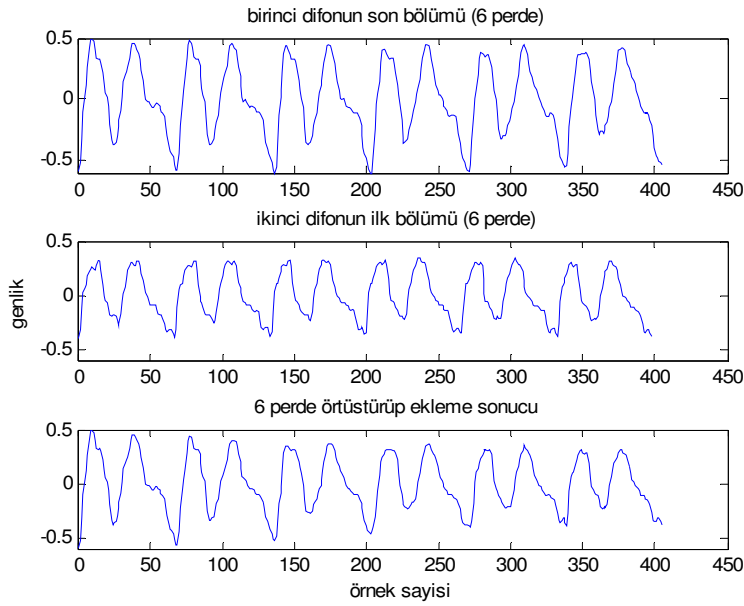
olarak alınmıştır. Aşağıda şekil 3.11’de, bir perde örtüştürüp ekleme sonucu; şekil 3.12’de, üç perde ve şekil 3.13’te de altı perde örtüştürüp ekleme sonucu gösterilmiştir. Ünlülerin birleştirilmesinde örtüştürme miktarı büyük (5-6 perde gibi) alınırken, ünsüz geçişlerinde örtüştürme miktarı küçük (1-2 perde) alınmaktadır.



Şekil 3.11 PSOLA yöntemiyle bir perde örtüştürüp ekleme sonucu



Şekil 3.12 PSOLA yöntemiyle üç perde örtüştürüp ekleme sonucu



Şekil 3.13 PSOLA yöntemiyle altı perde örtüştürüp ekleme sonucu

### 3.3.1 Perde işaretlerinin bulunması

Tez çalışmasında, kullanılan difonların perde işaretleri elle belirlenmiştir. Süre ve temel frekans değiştirme işlemleri sonrasında ise otomatik olarak güncellenmiştir (yeni yerleri hesaplanmıştır). Süre değiştirme işlemi sonucunda; süre uzatıldığında, tekrarlanan perdeler için perde işaretleri eklenmiş, süre kısaltıldığında ise, kullanılan perdelerin perde işaretleri kullanılmayanların üstüne yazılmıştır.

Perde noktalarının bulunması, difonların birleştirilmesinde ve süre ile temel frekans değişikliklerinde gereklidir. Çünkü PSOLA yönteminde işlemler, perde noktaları referans alınarak yapılır. Yapılan araştırmada görülmüştür ki; konuşma işlemede, perde işaretlerinin tam olarak belirlenmesi ve perde periyodu tahmini kesin olarak çözülebilmemiş problemler değildir. Bu çalışma kapsamında perde işaretleri, sinyal çerçevelerinin pozitif veya negatif tepe noktalarına konmaya çalışılmıştır.

Literatürde perde işaretlerinin bulunmasında şu yöntemler kullanılmaktadır:

- Teager enerji operatörü ile (Zhou vd. 2001),
- Dalgaşekli tepelerini kullanarak,
- Özilinti (autocorrelation) işlevini kullanarak,
- Enerji işlevini kullanarak (Goncharoff ve Gries 1998),

- **Teager Enerji Operatörü (TEO)**

Bu operatör, incelenen sinyalin doğrusal olmayan özelliklerini ortaya çıkarmaya yöneliktir.

Sürekli zamanda;

$$\Psi[x(t)] = \left( \frac{d}{dt} x(t) \right)^2 - x(t) \left( \frac{d^2}{dt^2} x(t) \right) \quad (3.7)$$

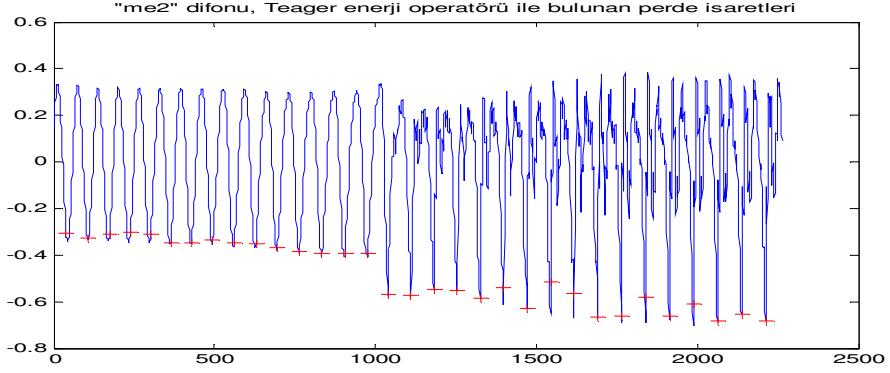
Eşitliği ile verilen işlem; ayrık zamanda:

$$\Psi[x(n)] = x^2(n) - x(n+1)x(n-1) \quad (3.8)$$

şeklindedir.

TEO ile perde işaretlerinin bulunmasında, öncelikle incelenen difonun Teager enerji işlevi hesaplanır. Daha sonra, 70 elemanlı çerçeveler boyunca, perde işaretleri bu işlevin maksimum olduğu noktalara yerleştirilir. 70 değeri;  $F_s = 16000$  Hz örnekleme frekansında işlenen kadın sesinin (veritabanında yer alan kayıtlar) ortalama periyodudur (örnek cinsinden).

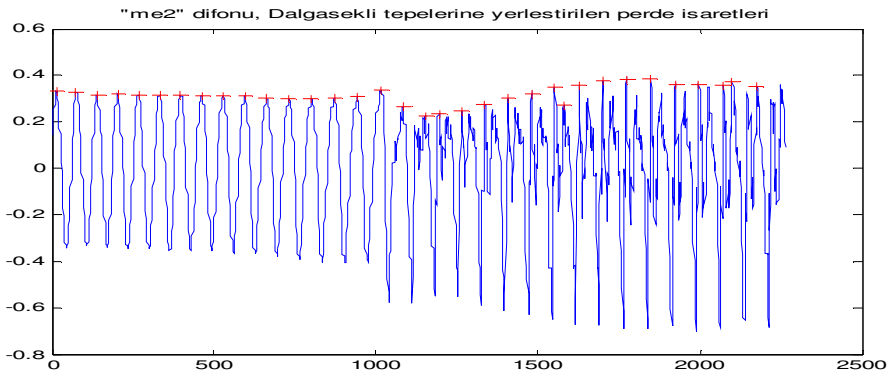
“m e2” difonunun TEO ile hesaplanan perde işaretleri şekil 3.14’te verilmiştir.



Şekil 3.14 “m e2” difonunda, TEO ile perde işaretlerinin bulunması

- **Dalgasekli Tepeleri ile Perde İşaretlerinin Bulunması**

Bu yöntemde, perde işaretleri ötümlü/ ötümsüz ayrımı yapıldıktan sonra, ötümsüz çerçevelerde 70 örnek arayla ( $F_s = 16$  kHz, kadın sesi için) düzenli olarak; ötümlü çerçevelerde ise sinyalin tepe noktalarına yerleştirilir. Bu yöntemde karşılaşılan sorun; perde katlarının [pitch doubling, halving] bulunabilmesidir. Şekil 3.15’te “m e2” difonunun dalgasekli tepeleri kullanılarak bulunan perde işaretleri gösterilmiştir.

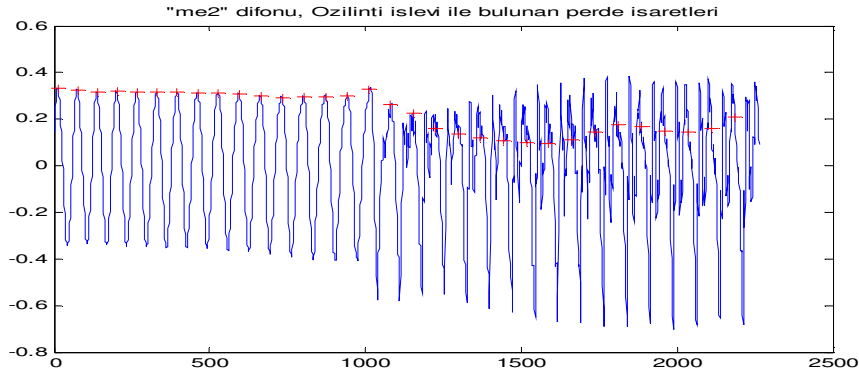


Şekil 3.15 “m e2” difonunda, dalgasekli tepeleri ile perde işaretlerinin bulunması

- **Özilinti İşlevi ile Perde İşaretlerinin Bulunması**

Bu yöntemde, ötümsüz çerçevelerde yine 70 örnek arayla yerleştirilen perde işaretleri, ötümlü çerçevelerde, 45-85 örnek gecikme için koşulan özilinti işlevi (eşitlik 3.10) sonucuna göre, en yüksek özilintiyi veren gecikme değerlerine yerleştirilir. “m e2” difonunun özilinti işlevi ile bulunan perde işaretleri şekil 3.16’da gösterilmektedir.

$$R_{xx}(l) = \sum_{n=45}^{85} x(n)x(n+l) \quad (3.9)$$

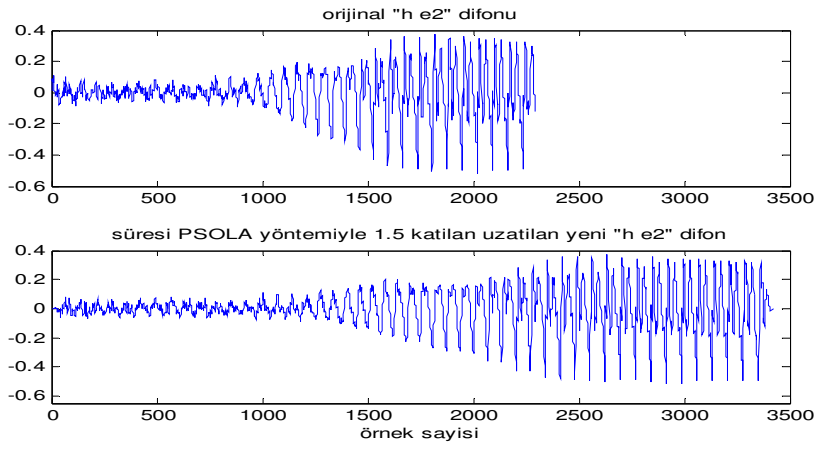


Şekil 3.16 “m e2” difonunda, özilinti işlevi ile perde işaretlerinin bulunması

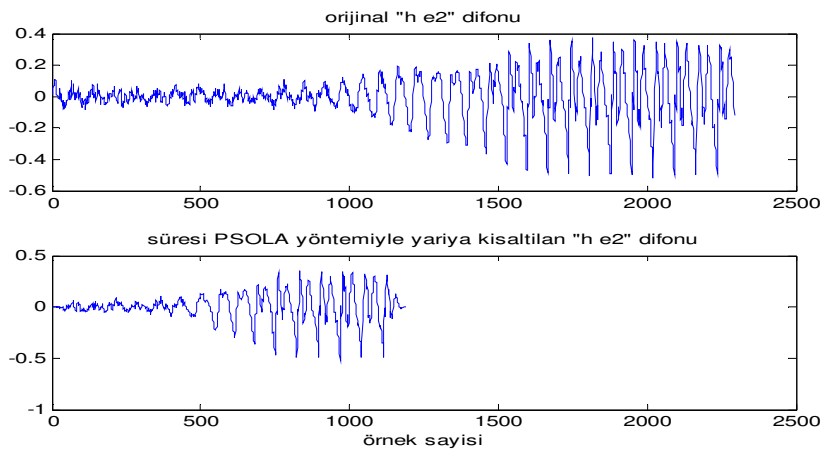
Bu üç yöntem karşılaştırılacak olursa, başarımlarının benzer olduğu ve üç yöntemin de bazı durumlarda perde işaretlerini hatalı bulduğu söylenebilir. Denenen tüm difonlarda, perde işaretlerini hatasız bulan yöntem olmadığı için, tez çalışmasında perde işaretleri elle belirlenmiş ve dosyalara kaydedilmiştir.

### 3.3.2 Süre değıştirme

Birleřtirilen ses paralarının uygun sürelerde olması için, ses paralarının uzatılması veya kısaltılması gerekir. Süre uzatmada, bazı perde periyotları tekrarlanır. Ötümlü perdeler tekrarlanırken, ötümsüzler tekrarlanmaz. Süre kısaltmada ise, bazı perde periyotları atılır. PSOLA ile gerekleřtirilen süre değıştirme işlemleri, řekil 3.17 ve řekil 3.18’ de görülebilir.



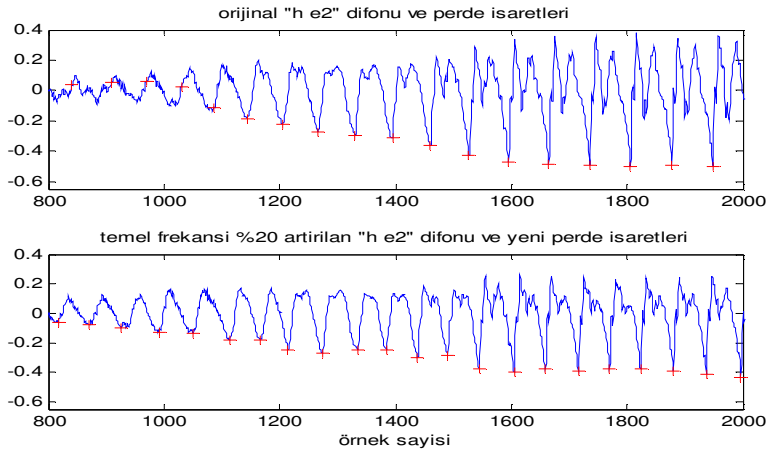
řekil 3.17 PSOLA ile süre uzatma



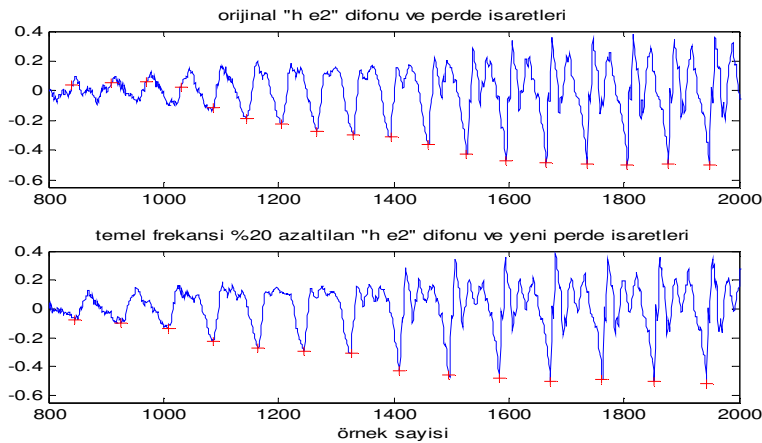
řekil 3.18 PSOLA ile süre kısaltma

### 3.3.3 Temel frekans deęiřtirme

Konuřma paralarının perde frekanslarının konuřmayı bozmadan deęiřtirilmesi nemlidir. Burada PSOLA yntemi (Trk 2000) kullanılmıřtır. Perde frekansı artırılırken, yeni perde iřaretleri birbirine yaklařacaktır. Bu durumda, kısalan srenin ara perdelerin tekrarlanması ile kompanse edilmesi gerekir. Perde frekansı azaltılırken, yeni perde iřaretleri birbirinden uzaklařacaktır. Bu durumda da, uzayan srenin ara perdelerin atılmasıyla kompanse edilmesi gerekir. PSOLA ile perde frekansının %20 artırılması řekil 3.20’de, perde frekansının %20 azaltılması ise řekil 3.21’de verilmiřtir.



řekil 3.19 “h e2” difonunun perde frekansının %20 artırılması ve yeni perde iřaretleri



řekil 3.20 “h e2” difonunun perde frekansının %20 azaltılması ve yeni perde iřaretleri

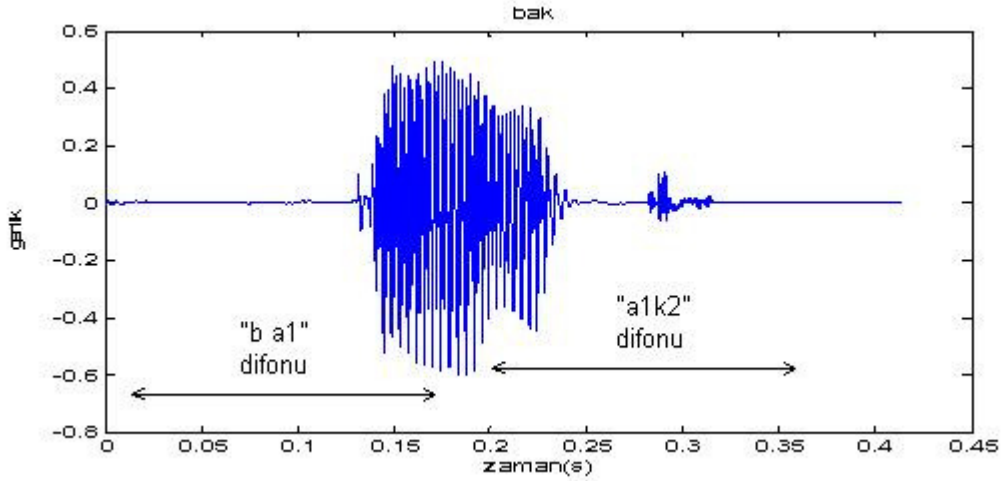
### 3.3.4 Enerji deęiřtirme

Enerji deęiřtirme; hazırlanan GUT'de en kolay yapılabilen iřlemdir. Konuřma parçasının enerjisinin deęiřtirilmesi için, tüm örnekler enerji oranının karekökü ile çarpılır (eřitlik (3.11)).  $E_1$ ; ilk difonun son  $K$  adet örneęinin enerjisi,  $E_2$  ise takip eden difonun ilk  $K$  adet örneęinin enerjisidir. Enerji oranının karekökü alınarak bulunan  $\alpha$  ile ikinci difon çarpılır ve enerjisi eřitlenen yeni difon elde edilir.

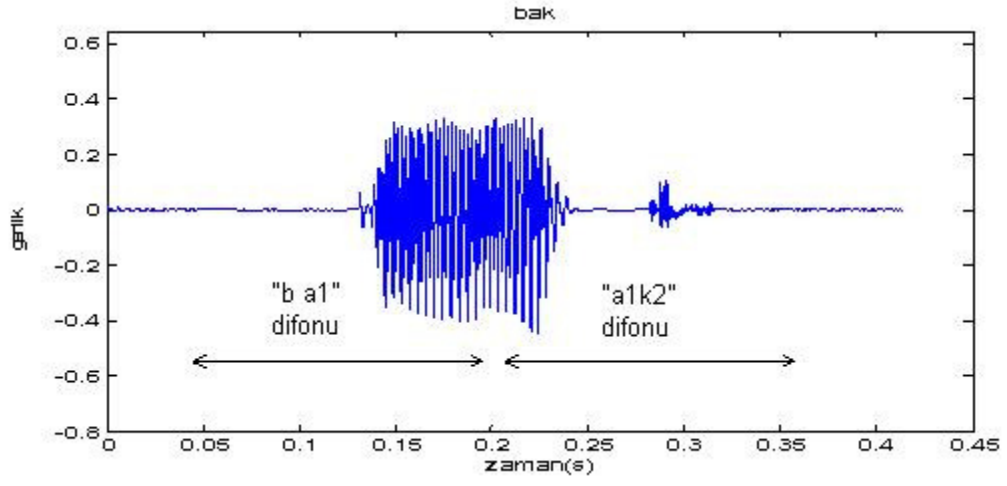
$$E_1 = \frac{1}{K} \sum_{n=1}^K d_1^2[end - n] \quad E_2 = \frac{1}{K} \sum_{n=1}^K d_2^2[n]$$
$$\alpha = \sqrt{\frac{E_1}{E_2}} \quad s_2 = \alpha \cdot d_2$$

(3.10)

Ařaęıda “b a1” ve “a1k2” difonlarının birleřtirilmesi ile elde edilen “bak” sentezinde enerji uyumu anlatılmıřtır.



řekil 3.21 Enerji eřitlenmemiř durumda sentez



Şekil 3.22 Enerji eşitleme sonrası sentez

#### 4. DOĞAL KONUŞMA İÇİN EZGİ MODELLEME

Bu bölümde tezin esas amacı olan doğal konuşma sentezi için önerilen ezgi modeli anlatılacaktır. Öncelikle prozodi, ezgi, vurgu ve aksan terimleri arasındaki fark ve bağlantıları belirtelim. En genişten daraya doğru gidersek; prozodi ve ezgi arasındaki fark, Mixdorff (1998) tarafından şöyle anlatılmaktadır: Prozodi, ezgiden daha geniştir ve ezgiye ek olarak, parçaların bağli süre ve enerjileri gibi başka parçalarıüstü özelliklerini veya ses kalitesi, ünlülerin formant frekansları gibi diğer özellikleri de inceler. Sözcük vurgusu; (Ergenç 2002)'de “accent” terimi ile ifade edilmiştir. Diğer kaynaklarda vurgu terimi “stress” ile belirtilmektedir (Şayli 2002, Kabak ve Vogel 2005, Öztürk 2005, Aydemir ve Yılmaz 2010). Bu iki terim aynıdır. Ezgiyi oluşturan akustik parametreler: süre, perde frekansı (temel frekans) ve enerjidir. Sözcük vurgusu ile Türkçenin ezgi modelleri daha önceki bazı çalışmalarda incelenmiştir. (Oskay vd. 2001), farklı yapılarıdaki (olumlu, olumsuz, soru cümleleri için) cümlelerin ezgisini; perde frekansını doğrusal ve ikinci dereceden fonksiyonlar kullanarak modellemeye çalışmışlardır. Şayli (2002), ezginin önemli bileşenlerinden biri olan süre parametresini incelemiştir. Fonem ve trifonların ortalama sürelerini hesaplamıştır. Trifon ağaç modeli ile toplamsal süre modeli Türkçe için geliştirilmiştir.

Fonem ortalama süreleri çizelge 4.1'de verilmiştir. Sentezleyicimizde difonlar kullanıldığı için, difon ortalama süreleri fonemlerin sürelerinin ortalaması alınarak hesaplanmıştır. Hatırlanacağı üzere, difon; ilk fonemin ortasından ikinci fonemin ortasına kadar olan bölüm şeklinde tanımlanmaktadır. Uzun fiil çekimlerini sözcük gibi değil, cümle gibi düşündüğümüzde, uzun sentezlerde Şayli'nin kelime ve cümle içi kullanım ayrımıındaki süre sıkıştırma değerleri burada da kullanılabilir.

Çizelge 4.1 (Şayli 2002)’den alınan ortalama fonem süreleri ve sıkıştırma oranları

| <b>Fonem</b> | <b>Sözcük içinde süre (ms)</b> | <b>Cümle içinde süre (ms)</b> | <b>Sıkıştırma oranı</b> |
|--------------|--------------------------------|-------------------------------|-------------------------|
| a            | 139                            | 112                           | 1.24                    |
| b            | 68                             | 55                            | 1.24                    |
| c            | 75                             | 67                            | 1.11                    |
| ç            | 118                            | 105                           | 1.13                    |
| d            | 58                             | 47                            | 1.23                    |
| e            | 135                            | 105                           | 1.29                    |
| f            | 99                             | 71                            | 1.40                    |
| g            | 64                             | 48                            | 1.33                    |
| h            | 67                             | 52                            | 1.29                    |
| ı            | 115                            | 81                            | 1.42                    |
| i            | 116                            | 82                            | 1.42                    |
| j            | 119                            | 73                            | 1.62                    |
| k            | 124                            | 83                            | 1.49                    |
| l            | 69                             | 56                            | 1.23                    |
| m            | 84                             | 72                            | 1.17                    |
| n            | 101                            | 72                            | 1.40                    |
| o            | 133                            | 109                           | 1.22                    |
| ö            | 133                            | 110                           | 1.22                    |
| p            | 105                            | 76                            | 1.38                    |
| r            | 75                             | 60                            | 1.24                    |
| s            | 134                            | 112                           | 1.20                    |
| ş            | 144                            | 123                           | 1.17                    |
| t            | 104                            | 79                            | 1.32                    |
| u            | 113                            | 81                            | 1.40                    |
| ü            | 109                            | 84                            | 1.30                    |
| v            | 66                             | 52                            | 1.27                    |
| y            | 64                             | 45                            | 1.42                    |
| z            | 109                            | 80                            | 1.36                    |

Amaç; tek bir nefesle söylenen uzun cümleleri doğru sürelerle elde etmektir. (Öztürk 2005)’te, fonem süreleri ve temel frekans eğrileri için istatistiksel modeller

geliştirmiştir. Hem süre, hem de perde frekansını modellerken; fonem türü, önceki ve sonraki fonemlerin özellikleri, seslem sayısı, seslemin sözcük içindeki konumu, vurgu alması–almaması gibi metinsel özelliklerin etkisini incelemiştir.

Görüldüğü üzere, daha önce yapılan çalışmalarda genel olarak izlenen yol; mümkün olduğunca fazla konuşmacıdan sözcük ve cümle kayıtları almak, bunları seslem (hece) türü, seslem sayısı, sözcük içerisindeki konum, ünlü-ünsüz sesbirim özellikleri vb. parametrelere göre modellemeye çalışmaktır. Arısoy vd. (2004), yaptıkları çalışmanın sonucunda, sözcüğün ilk sesleminde yer alan ünlülerin, sözcüğün son sesleminde yer alan ünlülere göre belirgin ölçüde kısa olduğunu, ayrıca geniş ünlülerin {a, e, o, ö}, dar ünlülere {ı, i, u, ü} göre daha uzun olduğunu rapor etmişlerdir.

Şayli'nin tezinde belirttiği sesbirim ortalama sürelerinden yararlanılarak difon ortalama sürelerine geçilmiş ve sistemin baştan bu sürelerle hazır olması sağlanmıştır (Şayli 2002).

Yine aynı çalışmadaki bulguların en önemlilerinden biri, seslem türünün süre üzerindeki etkisidir. Örnek olarak; ünlü ile biten hecelerde (“**ke**”lam, “**ta**”rak, “**ki**”tap gibi), ünlünün süresi kısadır. Yine, ünlüden sonra iki ünsüzün geldiği hecelerde (“**art**”çı, “**üst**”teki, k“**öş**k”), ünlünün süresi uzundur.

Şayli'nin verdiği sonuçlardan biri de Türkçe’de genel olarak seslem süresinin sözcük sonunda en uzun, ortada kısa ve başta en kısa olduğu bilgisidir.

Örnek:    “**na**”kit (“na” kısa),            ka“**na**”ma (“na” daha uzun),            bada“**na**” (en uzun)

Bir diğer bulgu da, sözcükte hece sayısı arttıkça hece sürelerinin azaldığıdır. Bu gözlem de; sözcük ve cümle sentezleri sırasında teyit edilmeye çalışılmıştır.

#### 4.1 MKS Sistemleri İçin Ezgi Modelleri

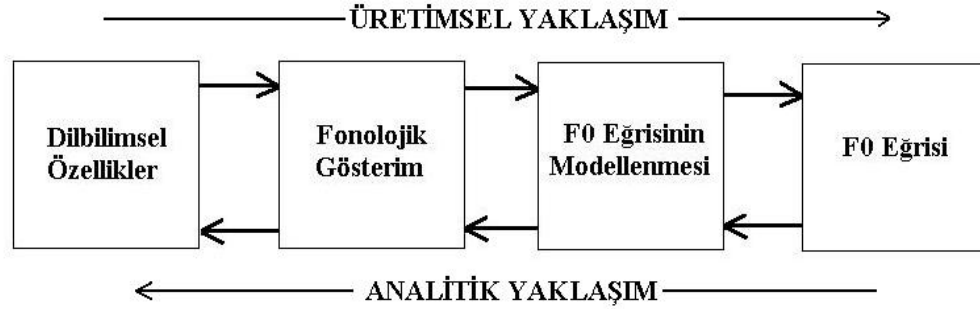
MKS sistemleri için pek çok ezgi modeli geliştirilmiştir. Genel olarak bu modelleri iki grup altında toplayabiliriz (Sproat 1998).

- sıralı ton modelleri (tone-sequence models)
- toplamsal modeller (superpositional models)

Sıralı ton modeli yaklaşımında;  $F_0$  eğrisinin, birbiriyle ilişkili olmayan farklı fonolojik ton sıralılarından oluştuğu düşünülmektedir. Burada ton ile kastedilen hece vurgusudur. Toplamsal model yaklaşımında ise;  $F_0$  eğrisi, lokal ve global etkili, birbiriyle hiyerarşik yapı içinde olan (tamlama ve aksan), aynı anda pek çok bileşenin toplamı sonucundaki karmaşık örüntü olarak düşünülmektedir. Yani; birinci grubu ayrık [discrete], ikinci grubu ise sürekli [continuous] modeller olarak düşünebiliriz. Fujisaki ezgi modeli (Fujisaki 1983) sürekli bir model iken; Pierrehumbert'in (Pierrehumbert 1980) ve Taylor'un (Taylor 1994) modelleri ayrık yapıdadır.

Ezgi modellerini gruplamada bir diğer sınıflandırma; 1- Akustik modeller, 2- Duyumsal [perceptual] modeller, 3- Dilbilimsel modeller, şeklindedir (Dutoit 1997). Bu sınıflandırmaya göre, Fujisaki ezgi modeli ve akustik stillendirme modelleri (Pierrehumbert 1980, Taylor 1994), akustik modeller sınıfına dâhildir. Duyumsal modeller sınıfında, IPO (Institute for Perception Research, Eindhoven) modeli ('t Hart vd. 1990) ile d'Alessandro ve Mertens'in (1995) geliştirdiği otomatik duyumsal ezgi modeli yer almaktadır. Dilbilimsel modeller ise, o dile özgü ezgi yapılarını kullanarak parçalarüstü [suprasegmental] birimler arasında uygun geçişler bulmayı amaçlar (Pierrehumbert 1980).

Mixdorff ise, ezgi modellerini, üretimsel [generative] yaklaşım ve analitik yaklaşım olarak iki gruba ayırmaktadır. (Mixdorff 1998). Üretimsel modellerde; metin özelliklerinden, dilbilimsel analiz ile,  $F_0$  eğrisine uygulanacak değişikliklere ve akustik çıkışa geçilir. Analitik modellerde ise, tam tersi olarak (şekil 4.1);  $F_0$  eğrisi, analitik inceleme sonucunda, matematiksel, grafiksel veya duyumsal olarak modellenmeye çalışılır ve buradan dilbilimsel özelliklere geçilir.



Şekil 4.1 Üretimsel ve analitik ezgi modeli yaklaşımı (Mixdorff 1998)

## 4.2 Geliştirilen Süre Modeli

Sentezleme sırasında, kayıtlı difonların süreleri nasıl ayarlanmalıdır? Gözlemlere göre; Türkçe sözcüklerde seslem süreleri, hecenin başta, ortada ve sonda oluşuna göre değişiklik göstermektedir. Genellikle, baştaki seslemler kısa, ortadakiler nispeten uzun ve sondaki seslemler de uzun sesletilir. Bunun temel sebebi; vurgunun genellikle son hecede bulunmasıdır. Dikkate alınması gereken ilk özelliğin bu olduğu düşünülmektedir.

Tez çalışmamızda; süre modelleme için kayıtlar üzerinden gitmek yerine, test-yatağında denemeler yapılması ve sentezlenen seslerin duyumsal olarak değerlendirilmesi tercih edilmiştir. Metinsel özellikler, difon süreleri ve örtüşme miktarları parametre olarak alınarak, kulağa hoş gelen sentezler için kurallar bulmaya çalışılmıştır [analysis by synthesis]. Bu amaçla denemelerimize, Ek 1’de anlatılan en çok rastlanan digram-difonlar ile başladık. Sentezlemede sözcükten cümleye geçtikten sonra, sözcükteki difon sayısının sürede etkili olduğunu fark ettik ve difon sayısına göre bir sistematik geliştirdik. Tasarlanan MKS sisteminin yaptığı sentezlerde aşağıda anlatılan difon sayısı kriteri dikkate alınmıştır (çizelge 4.2).

Çizelge 4.2 Difon süresi değiştirme parametresi ( $\alpha_d$ )

|                   |                                |
|-------------------|--------------------------------|
| $\alpha_d = 1,20$ | Difon sayısı $\leq 3$          |
| $\alpha_d = 1,00$ | $4 \leq$ Difon sayısı $\leq 6$ |
| $\alpha_d = 0,95$ | Difon sayısı $\geq 7$          |

Buna göre, sentezlenecek sözcükteki difon sayısı 4'ten az ise bu sözcüğün difonları %20 uzatılır ( $\alpha_d = 1,20$ ); eğer sözcükteki difon sayısı 6'dan fazla ise, difonların süresi %5 azaltılır ( $\alpha_d = 0,95$ ); eğer difon sayısı 4,5 veya 6 ise difonlar ortalama sürelerinde bırakılır ( $\alpha_d = 1,00$ ). Buradaki amaç; az difondan oluşan sözcüklerin daha uzun sentezlenmesi ve sentezin daha anlaşılır olmasıdır. Difon sayısı fazla olan sözcüklerde ise sentezleyici yavaş kalmaktadır. Bu da doğallığı olumsuz yönde etkilemektedir. Dolayısıyla difon sayısı çok olan sözcük sentezlerinde süre kısaltma uygulanmaktadır.

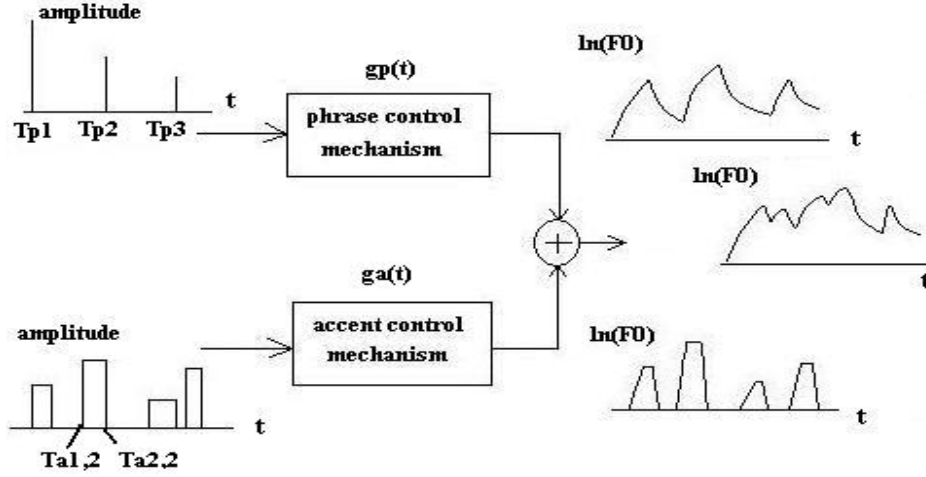
### 4.3 Fujisaki Ezgi Modeli

Fujisaki ezgi modeli toplamsal bir model olup, temel frekans değişimini, tamlama (phrase,  $P$ ) ve aksan (accent,  $A$ ) olmak üzere iki temel bileşenin toplamı şeklinde ifade eder. Modelin oldukça başarılı sonuçlar verdiği belirtilmektedir (Mixdorff 1998, Gutierrez-Arriola vd. 2001). Cümle ezgisini temsil eden perde frekansı eğrisi ( $F_0$ ), eşitlik (4.1) ile verilir.

$$\begin{aligned}
 \ln(F_0) &= \ln(F_{\min}) + P + A \\
 P &= \sum_{k=1}^{N_p} A_{p,k} \cdot g_p(t - T_{p,k}) \\
 A &= \sum_{k=1}^{N_a} A_{a,k} [g_a(t - T_{a1,k}) - g_a(t - T_{a2,k})]
 \end{aligned} \tag{4.1}$$

Burada,

$g_p(t) = \alpha^2 t \cdot e^{-(\alpha t)} u(t)$ ; tamlama dürtü tepkisi,  $g_a(t) = \min[1 - (1 + \beta t)e^{(-\beta t)}, \gamma]$ ; aksan basamak tepkisi,

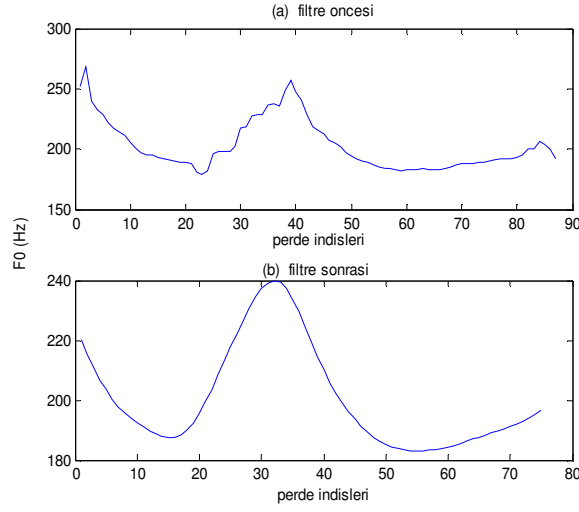


Şekil 4.2 Fujisaki ezgi modelinin blok şeması

$A_{p,k}$ ,  $A_{a,k}$ ,  $T_{p,k}$ ,  $T_{a1,k}$ ,  $T_{a2,k}$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ ; model parametreleri ve  $F_{\min}$ ; taban frekansıdır.

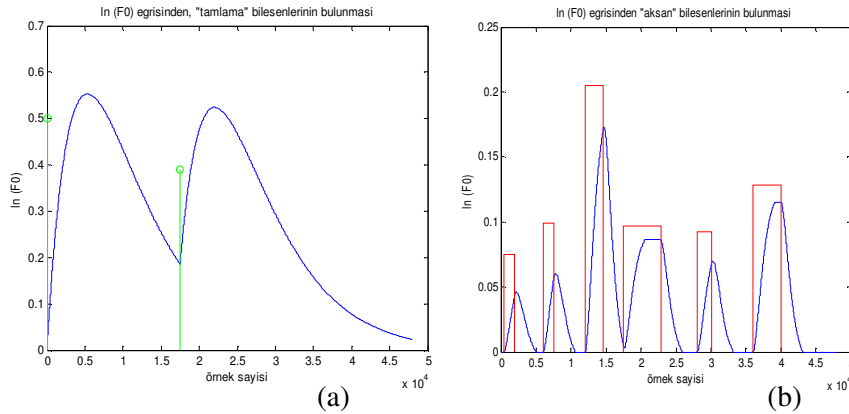
Bu modellemenin yapılabilmesi için önce  $F_0$  eğrisinin monoton bir forma getirilmesi gerekir. Şekil 4.3.a'da, konuşmacı tarafından seslendirilen “kokar” sözüne ait perde frekansı eğrisi görülmektedir. Bu eğride gözlenen ve mikro-bürünsel (micro-prosodic) etkilerden kaynaklanan ani değişimler, kayan-ortalamalı bir süzgeç yardımıyla giderilmiştir. Süzgeç çıkışı şekil 4.3.b'de görülebilir.

Bu çalışma kapsamında; pek çok dildeki MKS sistemlerinde perde frekansı değişiminin modellenmesinde kullanılan Fujisaki ezgi modelini Türkçedeki çekimli fiiller için uygulanmıştır.



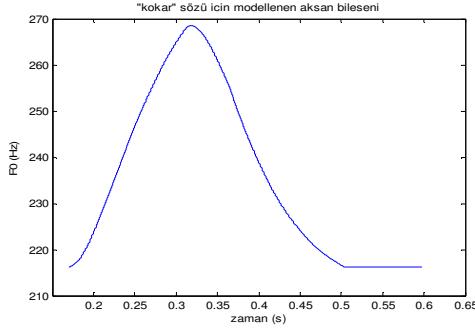
Şekil 4.3. (a,b) Kayan ortalamalı süzgeçle (N=7) “kokar” sözü  $F_0$  eğrisinin düzgünleştirilmesi

Model parametrelerinin bulunması için otomatik yöntemler kullanılır (Rossi vd. 2002). Bir cümle ezgisi için;  $F_0$  eğrisi, 1s süreli bölümler halinde incelenerek yerel maksimum noktaları bulunur. Bu tepe noktalarından önceki yerel minimum noktaları tamlama başlangıç noktalarını ( $T_{p,k}$ ) verir. Bu referans noktalarından başlamak üzere,  $g_p(t)$ : tamlama dürtü fonksiyonu oluşturulur. Dürtü genliği ( $A_{p,k}$ ); ilgili zaman aralığında  $F_0$  eğrisiyle hatayı minimize edecek şekilde yinelemeli olarak hesaplanır. Şekil 4.4’te, incelenen bir cümle için, tamlama (a) ve aksan (b) bileşenlerinin yerleri ve genlikleri görülmektedir.



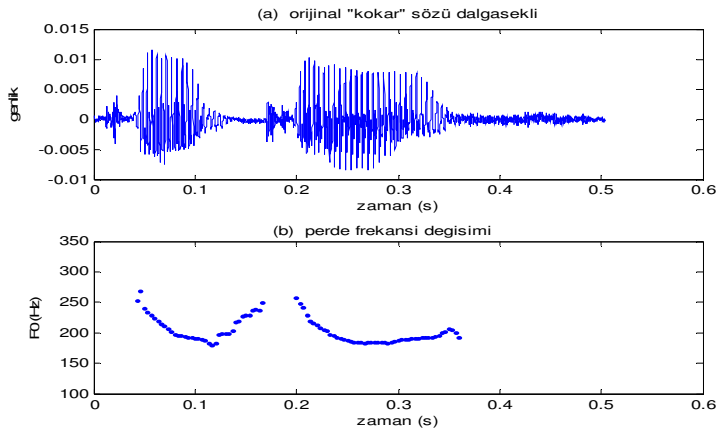
Şekil 4.4 a. Tamlama bileşenlerinin b. aksan bileşenlerinin yer ve genliklerinin bulunması

Hesaplanan tamlama bileşeni, modellenmek istenen  $F_0$  eğrisinden çıkartılır. Elde edilen fark sinyalindeki her tepe ayrı bir basamak fonksiyonu:  $g_a(t)$  ile ifade edilir. Bu işlem şekil 4.4.b’de gösterilmiştir.  $T_{a1,k}$  ve  $T_{a2,k}$  parametreleri; bu basamak fonksiyonlarının, sırasıyla, başlangıç ve bitiş zamanlarıdır. Şekil 4.5’te “kokar” sözünde hedeflenen ezgiyi gösteren aksan bileşeni görülmektedir.



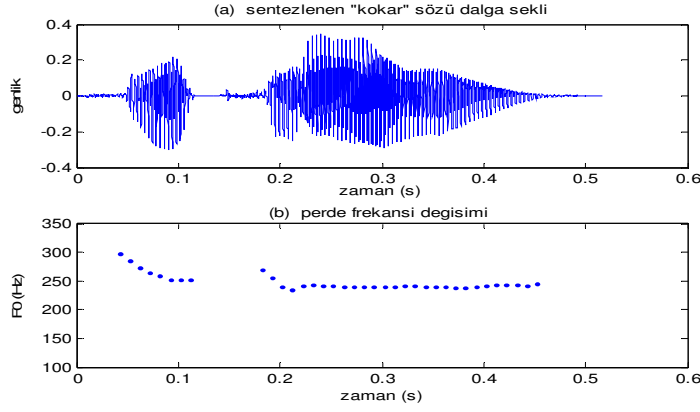
Şekil 4.5 “kokar” sözü için modellenen aksan bileşeni

Modellenen ezgi, daha doğal bir konuşma elde edebilmek için MKS sistemine uygulanmıştır. Bu amaçla sentezleme öncesinde, orijinal perde frekansları, model ile elde edilen frekans değerlerine değiştirilmiştir. Şekil 4.6.a’da özgün (ses kaydı) “kokar” sözü dalga şekli ve şekil 4.6.b’de  $F_0$  eğrisi verilmiştir.



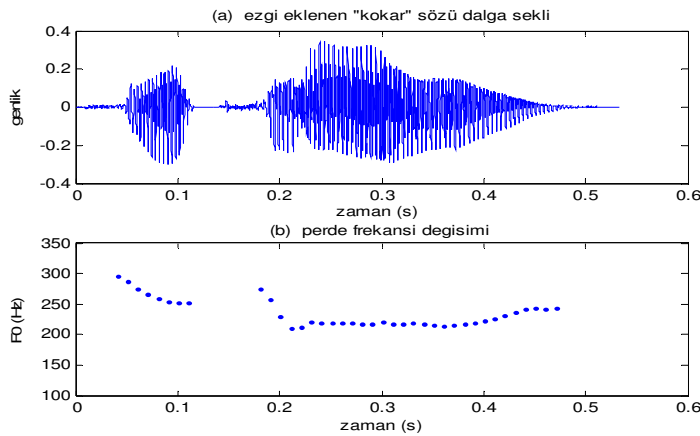
Şekil 4.6 Ses kaydı olan “kokar” sözünün, a. dalgaşekli, b.  $F_0$  eğrisi

Şekil 4.7.a’da, geliştirilen MKS sistemi çıkışının, var olan (kendisinden gelen) ezgiyle, dalga şekli ve şekil 4.7.b’de  $F_0$  eğrisi görülmektedir.



Şekil 4.7 Geliştirilen MKS sisteminin sentezlediği “kokar” sözünün, a. dalgaşekli, b.  $F_0$  eğrisi

Şekil 4.8.a’da ise, Fujisaki ezgi modeliyle ezgisi değiştirilen “kokar” sözü dalga şekli ve şekil 4.8.b’de  $F_0$  eğrisi görülmektedir. Sonucun doğal kaydın ezgisel özelliklerine daha yakın olduğu görülebilir.



Şekil 4.8 Ezgisi Fujisaki modeliyle değiştirilen “kokar” sözünün, a. dalgaşekli, b.  $F_0$  eğrisi

## 4.4 Geliştirilen Ezgi Modeli

Bu bölümde, Türkçe metinden konuşma sentezlemede doğal bir ezgi için yaptığımız modelleme anlatılacaktır.

### 4.4.1 Çekimli Fiillerde Vurgu

Türkçede fiil çekimleri, istisnalar haricinde, kurallara bağlıdır. Aydemir ve Yılmaz (2010), fiillerin çekimlenmesi ve vurgularının belirlenmesi üzerinde çalışmışlar, Türkçedeki yaklaşık 4600 adet fiilden 1100 adedi için 5400 farklı çekimin başarıyla yapıldığını ve vurgu konumlarının doğru tespit edildiğini belirtmişlerdir. Bu sonuçlar bizim de çalışmamızın temelini oluşturmaktadır.

Türkçe fiil çekimlerinde vurgunun yeri belirlidir. Biz de fiillere, olumlu, olumsuz, olumlu soru ve olumsuz soru yapıları için vurguyu en doğal şekilde vermeye çalıştık. Haber kipi ve dilek kipinde örnek fiiller seçip (haber kipi için: *gez-di-m*, *gez-me-di-m*, *gez-dim mi?*, *gez-me-dim mi?*, dilek kipi için: *kavra-malı-sın*, *kavra-ma-malı-sın*, *kavra-malı mı-sın?*, *kavra-ma-malı mı-sın?*), aşağıda Çizelge 4.3'te verilen değişiklikler ile sentezler yaptık.

Çizelge 4.3 Ezgi modelimizde uygulanan değişiklikler

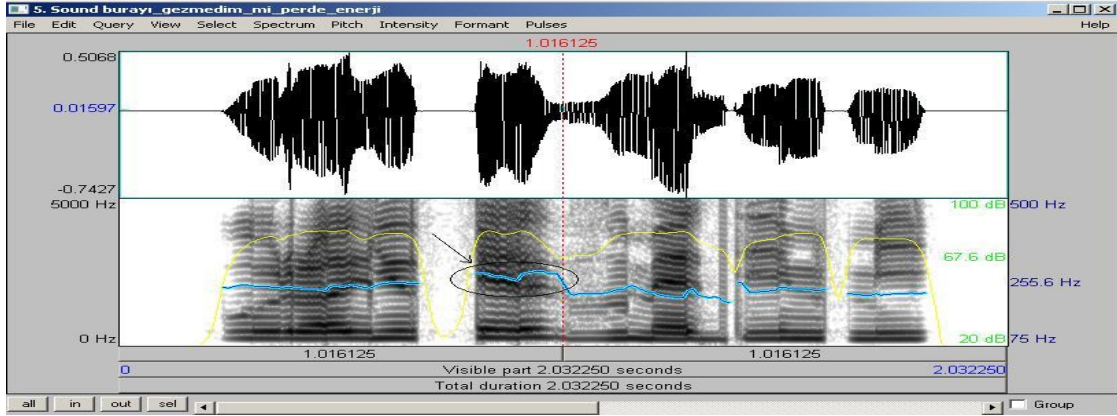
| Dosya Kodu | Uygulanan Değişiklik(ler)      |
|------------|--------------------------------|
| 1          | Ham sentez                     |
| 2          | Yalnızca süre                  |
| 3          | Yalnızca perde frekansı        |
| 4          | Yalnızca enerji                |
| 5          | Süre + perde frekansı          |
| 6          | Süre + enerji                  |
| 7          | Perde frekansı + enerji        |
| 8          | Süre + perde frekansı + enerji |

Süre değiştirmede, vurgu alan difonların süresi %40 artırılır.

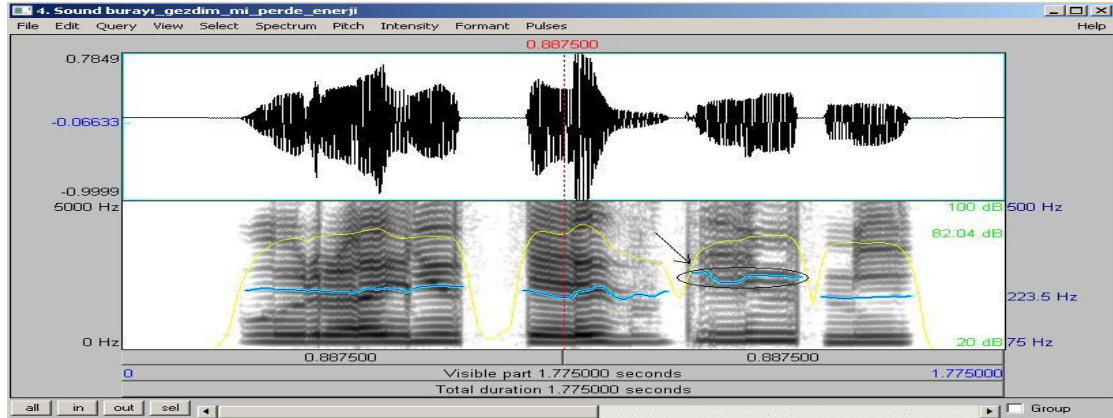
Perde frekansı değiştirmede, vurgu alan difonların perde frekansı %20 artırılır.

Enerji deęiřtirmede, vurgu alan difonların enerjisi %40 artırılır.

Bu oranlar, arayüzde yapılan pek çok denemenin ardından, dinleme testlerinin sonuçlarına göre bulunmuřtur.



řekil 4.9 “Burayı gezdim mi?” olumlu soru cümlesi (“-dim” sesleminde perde frekansı ve enerji deęiřtirme uygulanmıřtır)



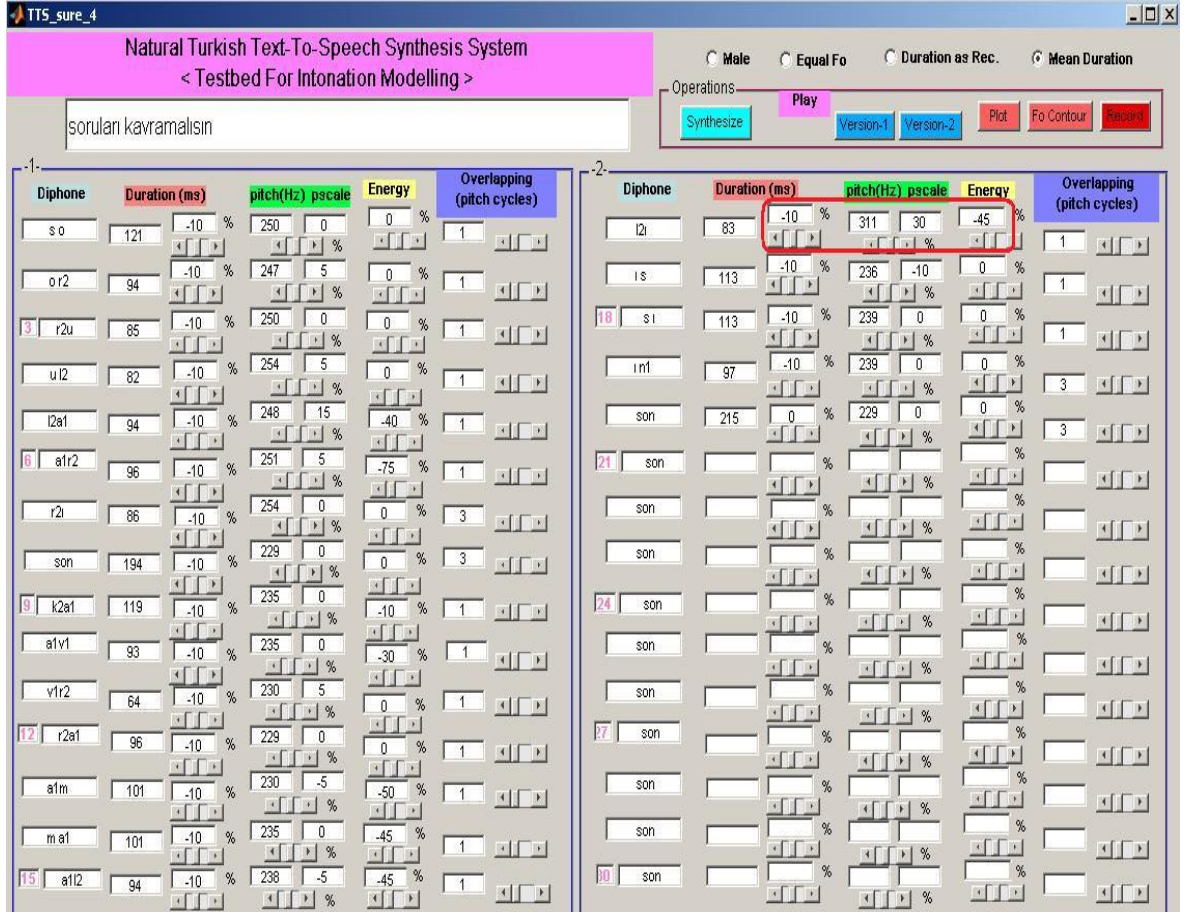
řekil 4.10 Burayı gezmedim mi?” olumsuz soru cümlesi (“gez” sesleminde perde frekansı ve enerji deęiřtirme uygulanmıřtır)

İncelenen cümlelerden bir tanesinin ham sentez ve perde frekansı deęiřtirme uygulanmıř halleri için kullanıcı arayüzü resimleri řekil 4.11 ve řekil 4.12 ile verilmiřtir. řekil 4.11’deki ham sentezde yapılan ilk iřlemler; difon sürelerinin ortalama süreye getirilmesinin yanı sıra, ardıřık difonların temel frekanslarının ve enerjilerinin mümkün olduęunca eřitlenmesidir.

Bir dildeki doğal konuşmanın kendine has bir ezgisi vardır. Bu; o dildeki sözcük ve cümleler arasındaki harmoni olarak tanımlanabilir. Ezginin temel işlevinin, konuşan ve dinleyen kişilerin daha kolay iletişim kurmalarını sağlamak olduğunu söyleyebiliriz. Vurgu; ezginin önemli bileşenlerinden biridir. “Seslemlerden birinin diğerlerine göre daha baskılı ve daha soluklu söylenmesi” (Ergenç 2002) olarak tanımlanan vurgu; Türkçe sözcüklerde genellikle son hecede (seslemde) yer alır. Fakat, Türkçe eklemeli bir dil olduğu için, sözcük sonlarına gelen eklerden ötürü, vurgunun yeri değişebilir. Sezer vurgu kuralına göre (Sezer 1983), Türkçede yer adlarında vurgu, başta veya ortadaki seslemde bulunmaktadır. Inkelas ve Orgun (2003) ile Kabak ve Vogel (2005), Sezer vurgu kuralının ve dilimizdeki diğer vurguların geçerliliğini araştırmışlardır. Yukarıda belirtilen kurala uymayan pek çok istisnai durumun olduğu rapor edilmiştir. Tüm bunlara rağmen, Türkçede fiil çekimi kuralları belirlidir (Aydemir ve Yılmaz 2010). Çalışmamızda fiiller üzerinde yoğunlaşmamızın temel sebebi de bu kuralların genelgeçer olmasıdır.



Şekil 4.11 “Soruları kavramalısın.” cümlesinin ham sentez hali için arayüzdeki ayarlar



Şekil 4.12 “Soruları kavramalısın.” cümlesinin perde frekansı değiştirilmiş hali için arayüzdeki ayarlar

Yukarıda anlatıldığı gibi oluşturulan sentez kümesi, aşağıda şekil 4.13 ve şekil 4.14’te gösterilen ağ sayfasına yüklenmiş ve dinleyicilerin CMOS (Comparative Mean Opinion Score) testine açılmıştır. CMOS testinin puan tablosu çizelge 4.4’te verilmiştir.

Çizelge 4.4 CMOS puanlarının karşılığı

| CMOS Puanı | Anlamı             |
|------------|--------------------|
| 1          | çok az fark var    |
| 2          | biraz fark var     |
| 3          | fark var           |
| 4          | belirgin fark var  |
| 5          | çok büyük fark var |

Yapılan testte, her defasında ham sentezle karşılaştırılan ezgili dosyalar rastgele sıra ile çalınmıştır (A ve B'nin sırası). Bunun amacı, dinleyicileri herhangi bir şekilde yönlendirmemek ve etkilememek içindir.

Welcome baran uslu (67)  
[Log out](#)

**Açıklamalar:**

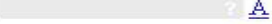
1. Lütfen, ses dosyalarının netliğinden ve kalitesinden çok, vurgunun doğruluğunu dikkate alınız.
2. A ve B dosyalarını birbiriyle karşılaştırınız. Hangisinde vurgulama daha başarılı, buna puan veriniz.
3. Dosyaları istediğiniz sırada, tekrar tekrar dinleyebilirsiniz.
4. Eğer mümkünse, testi kulaklık kullanarak yapınız.

Teşekkürler...

CMOS test

Şekil 4.13 Hazırlanan test sayfasında yer alan açıklamalar

1/32

    ? [A](#)

**A is ...**

☐ 5

☐ 4

☐ 3

☐ 2

☐ 1

☒ 0

☐ -1

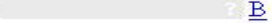
☐ -2

☐ -3

☐ -4

☐ -5

**... than B**

    ? [B](#)

Save

Şekil 4.14 Test sayfasından bir görünüm

#### 4.4.2 Sözcük ve öbek vurgusu

Çekimli fiillerde yaptığımız çalışmayı cümle geneline yayarak, aşağıda çizelge 4.5’te verilen cümleler için öbek vurgusu ile sözcük vurgularını senteze kazandırmaya çalıştık. Cümlede üzerinde durulan ve belirgin vurgu ile söylenen bölümler öbek vurgusunu oluşturur. Örnek olarak; “Her şeye rağmen zamanında geldi” cümlesinde “her şeye rağmen” söz grubu öbek vurgusunu taşımaktadır. Bu vurgunun senteze kazandırılması için, öbek vurgusunun yer aldığı difonların hem temel frekansları (%30), hem de enerjileri (%70) artırılır. Bu oranlara yine dinleme testleri sonucunda ulaşılmıştır.

Belirlenen ve incelenen cümleler çizelge 4.5’te verilmektedir.

Çizelge 4.5 Süre ve vurgunun birlikte incelendiği cümleler

| Cümle No | İncelenen cümle                         | Özellik                              |
|----------|-----------------------------------------|--------------------------------------|
| 1        | “Her şeye rağmen zamanında geldi.”      | Olumlu cümle                         |
| 2        | “Çok çalıştığı için başarılı oldu.”     |                                      |
| 3        | “Otobüsle uzun yola hiç gitmedim.”      | Olumsuz cümle                        |
| 4        | “Yıllardır güneş yüzü görmedi.”         |                                      |
| 5        | “Son sınava yeterince çalıştın mı?”     | Olumlu soru cümlesi                  |
| 6        | “Biz yokken kendine iyi baktın mı?”     |                                      |
| 7        | “Görevini en iyi şekilde yapmadın mı?”  | Olumsuz soru cümlesi                 |
| 8        | “Saçımı sana süpürge etmedim mi?”       |                                      |
| 9        | “Peki, yeterince çalışmıyor musun?”     | Kontrol cümlesi                      |
| 10       | “Noktalı virgül; açıklama içindir.”     | Noktalama işaretleri içeren cümleler |
| 11       | “En hakiki mürşit: ilimdir, fendir.”    |                                      |
| 12       | “Meyvelere örnekler: elma, armut, nar.” |                                      |
| 13       | “Çok yeteneklidir; ancak çalışmaz.”     |                                      |

Öncelikle, ortalama difon süreleriyle sentez yapılır. Birleştirmede bütün difonların perde frekansları ve enerjileri mümkün olduğu kadar eşitlenir. Bu senteze “*ham sentez*” adı verilir.

Bu cümlelere doğal ezgi kazandırmak için izlenen yol şu şekildedir:

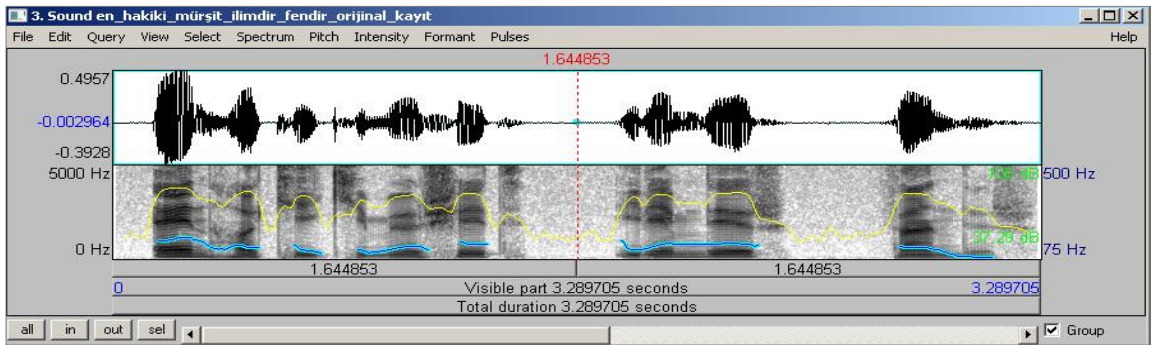
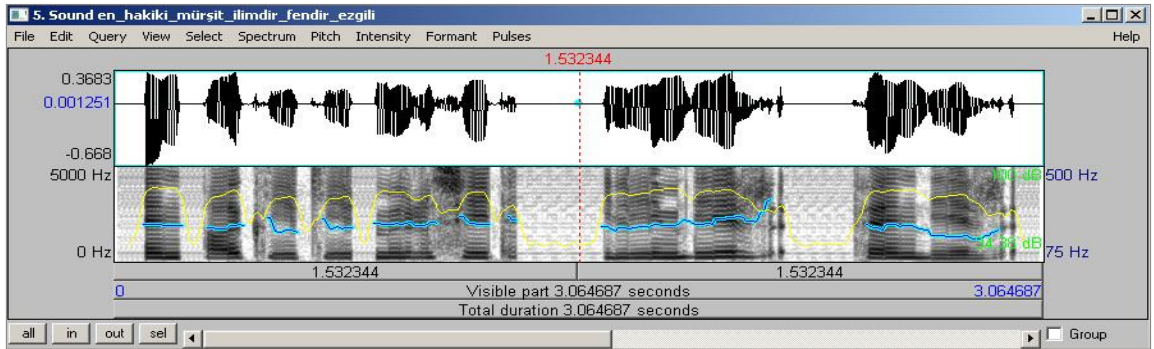
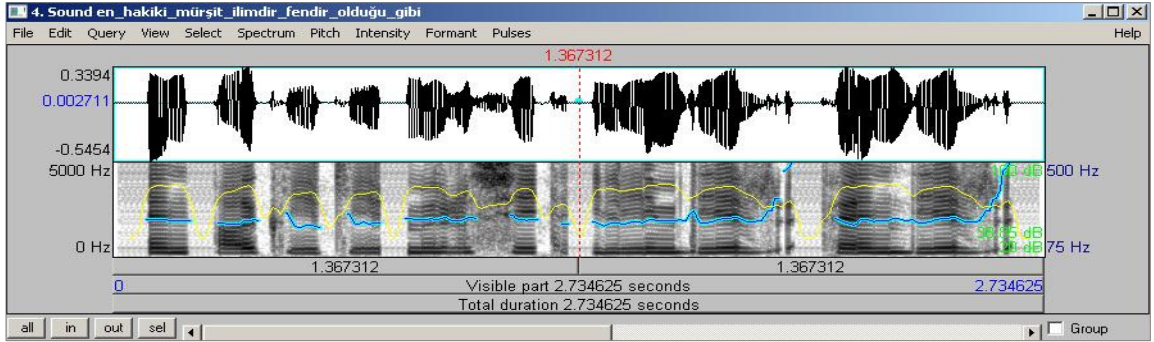
- Difon süreleri, çizelge 4.2’deki gibi ayarlanır.
- Söz gruplarının belirlediği duraklar boşluk süresi ile ayarlanır.
- Ham sentezi oluşturan tüm difonların temel frekansları ve enerjileri eşitlenmeye çalışılır.
- Orijinal kayıtlarda ortak olan vurgular senteze verilmeye çalışılır. Bunun için, öbek vurgusunun yer aldığı difonların hem perde frekansları (%30), hem de enerjileri (%70) arttırılır.
- Sözcük vurgularında ise sadece temel frekans %10 arttırılır.
- Cümlelerin fiiline vurgu katılır. Bu amaçla, fiilin vurgulu hecesindeki difonların temel frekansları %20, enerjileri %40 arttırılır (Uslu vd. 2011).
- Ayrıca cümle biten ezgi ile sonlandırılır. Bunun için, son heceyi oluşturan difonların temel frekansları ve enerjileri kademeli olarak (sırasıyla, %10, %15, %20) azaltılır.

Yukarıda anlatılan işlemler için geliştirilen ara yüzde yapılan ayarlamalara bir örnek şekil 4.15’te gösterilmiştir. “*Otobüsle uzun yola hiç gitmedim*” cümlesi için vurgu kurallarının uygulandığı bölümler yuvarlak içine alınmıştır. Burada öbek vurgusunun yer aldığı difonların (“*hiç*”) hem perde frekansları (%30), hem de enerjileri (%70) arttırılır (şekil 4.15’te mavi işaretli bölüm). Sözcük vurgularında ise sadece temel frekans %20 arttırılır (şekil 4.15’te kırmızı işaretli bölüm). Cümlelerin fiiline vurgu katılır. Bu amaçla, fiilin vurgulu hecesindeki difonların temel frekansları %20, enerjileri %40 arttırılır. Ayrıca cümle biten ezgi ile sonlandırılır. Bunun için, son sözcükteki difonların temel frekansları ve enerjileri kademeli olarak (sırasıyla, %10, %15, %20) azaltılır (şekil 4.15’te yeşil işaretli bölüm).



Şekil 4.15 “Otobüsle uzun yola hiç gitmedim” cümlesi için ara yüzdeki ayarlar

Şekil 4.16.a’da, “*En hakiki mürşit: ilimdir, fendir*” cümlesinin eşit temel frekanslı ve enerjili, ezgi içermeyen ham sentezi görülmektedir. Şekil 4.16.b’de ise söz gruplarının (*en hakiki mürşit*), fiil çekiminin (*fendir* – olumlu) ve noktalama işaretlerinin ezgili sentez hâli görülmektedir. Şekil 4.16.c’de ise, referans alınan orijinal konuşma kaydı görülmektedir. Ezgili sentezin orijinal kayda benzerliği; dalga şeklinin altında yer alan paneldeki sarı eğri ile gösterilen enerji konturundan (üstteki eğri) ve mavi eğri ile gösterilen temel frekans konturundan (alttaki eğri) fark edilebilir.



Şekil 4.16.a “En hakiki mürşit: ilimdir, fendir.” cümlesinin ham sentezi

Şekil 4.16.b “En hakiki mürşit: ilimdir, fendir.” cümlesinin ezgili sentezi

Şekil 4.16.c Orijinal kayıt: “En hakiki mürşit: ilimdir, fendir.”

Bu çalışmada, (Salor 2002)’de oluşturulan 16 kHz ile örneklenmiş difonlar kullanılmıştır. Ayrıca yapılan tüm sentezler, GoldWave programında, uzunluğu 3 olan yumuşatma (smoother) süzgecinden geçirilmiştir. Bu süzgecin amacı; tıslama ve çatırtı seslerinin etkisini azaltmaktır. Yapılan tüm sentezler,

<http://demo.reformo.net/baran4/index.php> 'de yer alan ağ sayfasındaki testten dinlenebilir.

#### 4.4.3 Noktalama işaretleri

Bilindiği gibi noktalama işaretleri, konuşmacıya duraklama süreleriyle ilgili bilgi vermek ve metnin doğru anlaşılması için kullanılır. Konuşurken veya okurken nefesimizi bu işaretlere göre ayarlarız. Sentezleyici metin içerisinde noktalama işaretleriyle karşılaşınca ne yapmalıdır? Aşağıda bu sorunun yanıtı için önerilerimiz yer almaktadır. Burada yalnızca duraklama amaçlı işaretler incelenmiştir. Tarihlerde ve kısaltmalarda nokta kullanılması gibi diğer noktalama işareti kullanımı ele alınmamıştır.

Noktalama işaretlerinin etkilerinin belirlenmesi için, bir televizyon kanalının yaklaşık 45 dakikalık haber bültenleri incelenmiştir (Kanal B, 19.09.2011 ve 08.10.2011 tarihli haber bültenleri). Metinde yer alan noktalama işaretleri, ses kayıtlarından tek tek elle bulunmuş ve ortalama durak süreleri hesaplanmıştır. Virgül, noktalı virgül ve iki nokta üst üste işaretleri incelenmiş, bulunan sonuçlar çizelge 4.6'da gösterilmiştir.

Çizelge 4.6 Noktalama işaretlerinin hesaplanan ortalama süreleri

| Noktalama işareti  | Ortalama süre |
|--------------------|---------------|
| virgül             | 220 ms        |
| noktalı virgül     | 260 ms        |
| iki nokta üst üste | 310 ms        |

- **Virgül**

Bilindiği gibi virgül, sıralanan denk sözcük, grup ve cümleleri ayırmak ve zarurî duraklar için kullanılır (Ergin 2002). Sentezleyicide, virgül olan yerlerde duraklama için, boşluk difonunun süresi normal sürenin %120 fazlası olarak alınmıştır.

- **Noktalı virgöl**

Virgülden daha belirgin ayırma ve duraklama işaretidir. Virgül ile nokta arası kuvvetli durak ve ayırmalarda kullanılır (Ergin 2002). Bu işaretin olduğu yerlerde, duraklama daha uzun tutulmuş, boşluk difonunun süresi, normal sürenin %160 fazlası olarak alınmıştır.

- **İki nokta üst üste**

Gelecek söze dikkati çeker. Açıklama yapmak veya örnek vermek istediğimiz zaman kullanılır (Banguoğlu 2007). İki nokta üst üstden sonra, duraklama en uzun tutulmuş, boşluk difonunun süresi, normal sürenin %210 fazlası olarak alınmıştır.

Özetle bu bölümde, tez çalışmamızın esası olan ezgi modeli yaklaşımlarımız anlatılmıştır. Takip eden bölümde ise modelin başarımı ve bulgular anlatılacaktır.

## 5. BULGULAR

Bu bölümde, 4. bölümde anlatılan ezgi modelinin başarımı ölçülmektedir. Yapılan dinleme testlerinden elde edilen bulgular bu bölümde anlatılmaktadır.

Yaşları 24-44 arasında değişen, herhangi bir duyum problemi olmayan ve anadili Türkçe olan 30 dinleyici, iki ayrı web sayfası üzerinden dinleme testlerimize katılmıştır. Testte yer alan (kontrol amaçlı konan) birbiriyle tıpa tıp aynı olan dört adet dosyaya, mutlak değerce 3 veya 3'ten daha fazla puan vermiş olan dinleyicilerin sonuçları dikkate alınmamıştır. Haber kipinde, “gez-mek” fiilinin di’ li geçmiş zamanda, Birinci tekil şahsa göre çekimlenmiş olumlu, olumsuz, olumlu soru ve olumsuz soru formlarına dinleyicilerin verdiği ortalama sonuçlar çizelge 5.1’de yer almaktadır.

Çizelge 5.1 “Burayı gezdim” cümlesinin olumlu, olumsuz, olumlu soru ve olumsuz soru formları için yapılan sentezlere dinleyicilerin verdiği ortalama puanlar (dosya kodu çizelge 4.3’te tanımlanmıştır)

| Dosya Kodu<br>Seçilen cümle             | 1     | 2     | 3           | 4    | 5     | 6     | 7           | 8    |
|-----------------------------------------|-------|-------|-------------|------|-------|-------|-------------|------|
| Burayı gez-’ <b>dim</b>                 | 0.03  | -0.3  | -1.7        | 0.1  | -2.23 | -0.57 | -1.43       | -2   |
| Burayı ‘ <b>gez</b> -me-dim             | -0.4  | -0.87 | -0.93       | 0.33 | -0.63 | -0.6  | -0.53       | -0.7 |
| Burayı gez-’ <b>dim</b> mi?             | 0.07  | 0.07  | <b>1.3</b>  | 0.13 | 0.5   | -0.13 | <b>1.37</b> | 0.73 |
| Burayı ‘ <b>gez</b> -me- <b>dim</b> mi? | -0.03 | -0.8  | <b>1.03</b> | 0.06 | -0.53 | -0.6  | <b>0.83</b> | 0.33 |

Dilek kipinde, “kavra-mak” fiilinin “malı” gereklilik kipinde, ikinci tekil şahsa göre çekimlenmiş olumlu, olumsuz, olumlu soru ve olumsuz soru formlarına dinleyicilerin verdiği ortalama sonuçlar çizelge 5.2’de yer almaktadır.

Çizelge 5.2 “Soruları kavramalısın” cümlesinin olumlu, olumsuz, olumlu soru ve olumsuz soru formları için yapılan sentezlere dinleyicilerin verdiği ortalama puanlar (dosya kodu Çizelge 4.3’te tanımlanmıştır)

| <b>Dosya Kodu</b>                 | <b>1</b> | <b>2</b> | <b>3</b>    | <b>4</b> | <b>5</b>    | <b>6</b>    | <b>7</b>    | <b>8</b> |
|-----------------------------------|----------|----------|-------------|----------|-------------|-------------|-------------|----------|
| <b>Seçilen cümle</b>              |          |          |             |          |             |             |             |          |
| Soruları kav-ra-ma-’lı-sın        | 0.07     | -0.21    | -0.71       | 0.29     | -1.5        | <b>0.86</b> | -0.07       | -1.21    |
| Soruları kav-’ra-ma-ma-lı-sın     | 0.07     | 0.29     | 0.36        | 0.07     | 0.43        | 0.07        | 0.43        | 0.07     |
| Soruları kav-ra-ma-’lı mı-sın?    | 0.14     | 0.36     | <b>0.86</b> | 0.43     | <b>1.43</b> | -0.43       | <b>1.57</b> | 0.5      |
| Soruları kav-’ra-ma-ma-lı mı-sın? | 0.36     | -0.36    | -0.43       | -0.36    | 0.5         | 0           | -0.07       | 0.43     |

Sonuç çizelgelerinde negatif değerler yer almaktadır. Bunun anlamı; yapılan değişikliğin doğallığı veya kaliteyi bozacak etkide bulunmasıdır. Ya vurgu kuralı yanlış, ya da vurgu verilirken yanlış bir şey yapılıyor demektir. Olumlu ve olumsuz formlar için, perde frekansının içinde bulunduğu tüm kombinasyonlarda (3, 5, 7 ve 8) büyük negatif değerler elde edilmiştir. Demek ki, %20 perde frekansı değiştirme burada doğallığı bozmaktadır. Yapılmaması tercih edilmelidir. Bu formlar için ham sentez tercih edilmelidir.

Yine sonuçlarda sıfıra çok yakın olan değerler var. Bu değerlerin anlamı; o parametre pek işe yaramıyor demektir. Tüm çekim formlarında, süre ve enerji değişikliklerinin (sırasıyla 2 ve 4) aldığı puanlar genel olarak böyledir. Demek ki, süre ve enerji değiştirmenin ezgide etkisi fazla değildir ve bunlar temel frekansa göre ikincil öneme sahiptir. Bu sonuç, önceki çalışmalarla tutarlıdır. (Dutoit 1997, Mixdorff 1998)

Olumlu soru formunda, perde frekansı ve enerjinin birinci fiilde ortalama 1,37/5,00, ikinci fiilde de ortalama 1,57/5,00 puanlık iyileşme sağladığı görülmüştür. Bununla birlikte, birinci fiilde yalnızca temel frekans ortalama 1,30/5,00 puan iyileşme

sağlamışken, ikinci fiilde süre ve temel frekansın birlikte ortalama 1,43/5,00 puan iyileşme sağlamıştır. Bu tür cümlelerde, iyileşmedeki esas katkının temel frekanstan kaynaklandığı değerlendirilmektedir.

Olumsuz soru formunda, birinci fiilde yalnızca temel frekans ortalama 1,03/5,00; temel frekans ve enerji birlikte ortalama 0,83/5,00 puanlık iyileşme sağlamışsa da benzer iyileşme ikinci fiilde elde edilmemiştir. Bu sebeple olumsuz soru formları için de ham sentezin kullanılması önerilebilir.

Türkçe’de fiil çekimlerinin, kural tabanlı yaklaşımlar ile gerçekleştirilmesi mümkündür. Söz konusu kurallar, sadece fiil çekimi esnasındaki sözdizimi ile sınırlı olmayıp, vurgunun konumu için de geçerlidir.

Fiil çekimlerinden sonra, cümle bazında sentezler yapılmıştır. Yukarıdaki bilgiler ışığında yapılan cümle sentezleri CMOS testine tâbi tutularak, belirlediğimiz ezgi kurallarının doğallığa olan etkisi ölçülmüştür. Toplam 13 adet test cümlesine verilen ortalama CMOS puanları çizelge 5.3’te verilmiştir.

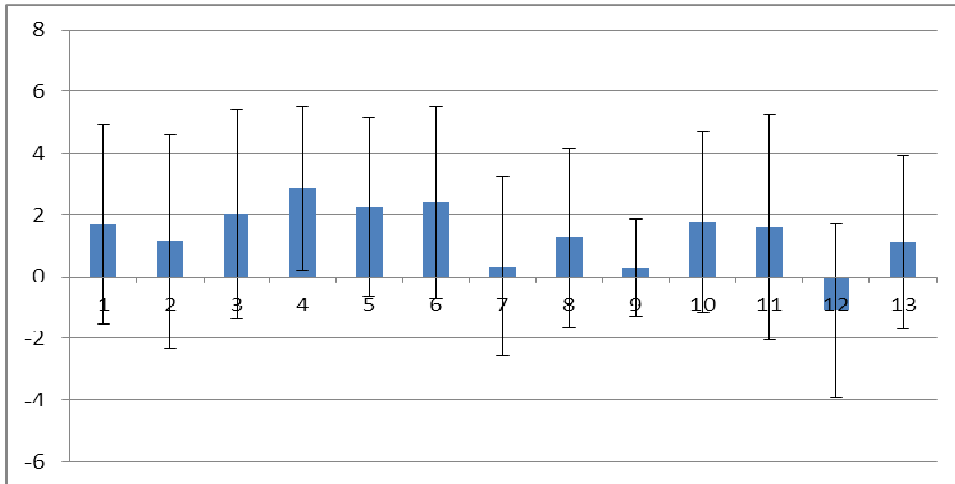
Çizelge 5.3 Test cümlelerine verilen ortalama CMOS puanları

| Cümle No | İncelenen cümle                     | CMOS Puanı (/ 5,0) |
|----------|-------------------------------------|--------------------|
| 1        | “Her şeye rağmen zamanında geldi.”  | 1,71               |
| 2        | “Çok çalıştığı için başarılı oldu.” | 1,14               |
| 3        | “Otobüsle uzun yola hiç gitmedim.”  | 2,00               |
| 4        | “Yıllardır güneş yüzü görmedi.”     | 2,97               |
| 5        | “Son sınava yeterince çalıştın mı?” | 2,11               |
| 6        | “Biz yokken kendine iyi baktın mı?” | 2,42               |

Çizelge 5.3 Test cümlelerine verilen ortalama CMOS puanları (devam)

|    |                                         |       |
|----|-----------------------------------------|-------|
| 7  | “Görevini en iyi şekilde yapmadın mı?”  | 0,33  |
| 8  | “Saçımı sana süpürge etmedim mi?”       | 1,29  |
| 9  | “Peki, yeterince çalışıyor musun?”      | 0,28  |
| 10 | “Noktalı virgül; açıklama içindir.”     | 1,77  |
| 11 | “En hakiki mürşit: ilimdir, fendir.”    | 1,60  |
| 12 | “Meyvelere örnekler: elma, armut, nar.” | -1,10 |
| 13 | “Çok yeteneklidir; ancak çalışmaz.”     | 1,13  |

Teste katılan ve kontrol cümlesine (9 numaralı cümle) +2, +1, 0, -1, -2 puanlarından birini veren 30 dinleyicinin ortalama CMOS puanları mutlak değerce 5’e normalize edildikten sonra şekil 5.1’de çizdirilmiştir. Bu bulgulara göre, 12 numaralı cümle hariç tüm ezgili sentezlerin ham sentezlerden daha doğal ve başarılı bulunduğu söylenebilir. En yüksek puanı 4 numaralı “Yıllardır güneş yüzü görmedi” cümlesi almıştır (ort. 2,97/5,00). Daha sonra en yüksek puanı 6 numaralı “Biz yokken kendine iyi baktın mı?” olumlu soru cümlesi almıştır (ort. 2,42/5,00). 12 numaralı “Meyvelere örnekler: elma, armut, nar.” cümlesinde yapılan değişiklikler, dinleyicilerin çoğunluğu tarafından beğenilmemiş olsa gerek ki negatif puan almıştır (ort. -1,10/5,00).



Şekil 5.1 Sentezlerin aldığı ortalama CMOS puanları ve  $\pm 1\sigma$  hata çubukları

Bu çalışmada Türkçe metinden konuşma sentezleme için yeni bir ezgi modeli önerilmiştir. Bu model; öbek vurgusunu, çekimli fiillerin vurgularını ve bazı noktalama işaretlerini dikkate almaktadır. Önerilen duyumsal ezgi modelinin başarımı öznel dinleme testleriyle değerlendirilmiş, CMOS testi sonucuna göre, ezgili sentezlerin ham sentezlerden ortalama 1,45/5,00 puan daha başarılı bulunduğu sonucu elde edilmiştir. Ezgi modelimiz sayesinde ortalamada, olumlu cümlelerde 1,43/5,00; olumsuz cümlelerde 2,49/5,00; olumlu soru cümlelerinde 2,27/5,00 ve olumsuz soru cümlelerinde 0,81/5,00 puanlık iyileşme sağlanmıştır. Noktalama işaretlerinin dikkate alındığı sentezlerde ise ortalama 0,83/5,00 puanlık iyileşme kaydedilmiştir.

Şekil 5.1’de ortalama CMOS puanları ve bir sigma (standart sapma) değerleri çizdirilmiştir.

Bu sonuçlara göre, temel frekansın ezgi için en önemli parametre olduğu, enerji ve sürenin daha az etkili oldukları, önerilen modelle Türkçe cümlelerde en doğal ezgiyi olumlu soru cümlelerinde elde edebileceğimiz söylenebilir.

## 6. SONUÇLAR VE TARTIŞMA

Bu tezde, Türkçe metinden konuşma sentezleme sistemleri için yeni bir ezgi modeli önerilmiştir. Çekimli fiiller üzerinde yapılan çalışmaların sonunda, aşağıdaki sonuçlar elde edilmiştir:

- Olumlu ve olumsuz cümle sentezlerinde, yapılan değişikliklerin büyük oranda kaliteyi düşürecek şekilde etki yaptığı, kalite artışı sağlayan değişikliklerin de kayda değer bir katkısının olmadığı gözlemlenmiştir. Dolayısıyla, bu tür cümlelerde, sentezleme işleminde bu çalışma kapsamında yapılmış olan tarzdaki değişiklikler yerine ham sentezin kullanılmasının daha uygun olacağı değerlendirilmektedir.
- Olumlu soru formunda, perde frekansının ortalama 1,37/5,00; perde frekansı ve enerjinin ortalama 1,51/5,00 puanlık iyileşme sağladığı görülmüştür. Bu tür cümlelerde, değişiklikteki esas katkının perde frekansından kaynaklandığı değerlendirilmektedir. Dinleme testi puanlarının dağılımı, sonuçların istatistiksel olarak anlamlı olduğunu göstermektedir.
- Olumsuz soru formunda, perde frekansının ortalama 1,38/5,00; perde frekansı ve enerjinin ortalama 0,79/5,00 puanlık iyileşme sağladığı görülmüştür. Benzer şekilde, bu tür cümlelerde de değişiklikteki esas katkının perde frekansından kaynaklanmakta olduğu değerlendirilmektedir. Yine dinleme testi puanlarının dağılımı, sonuçların istatistiksel olarak anlamlı olduğunu göstermektedir.

Türkçe’de fiil çekimlerinin, kural tabanlı yaklaşımlar ile gerçekleştirilmesi mümkündür. Söz konusu kurallar, sadece fiil çekimi esnasındaki sözdizimi ile sınırlı olmayıp, vurgunun konumu için de geçerlidir. Bu kurallar MKS sistemlerinde de uygulanabilir. Çalışmamızda bu düşünceyle hareket edilmiş ve özellikle soru cümlelerinde yapılacak olan değişiklikler ile sentez başarımının artırılacağı gözlemlenmiştir. Sonuçta MKS sistemlerinde fiil çekimleri için yeni ezgi kuralları tanımlanmıştır. Elde edilen sonuçlara göre, özetle; soru cümlelerinde vurgunun; vurgulu hecedeki difonların perde frekansının %20, enerjilerinin ise %40 artırıldığı bir ezgi modeliyle verilebileceği değerlendirilmiştir.

Fiil çekimlerinde uygulanan kurallar cümle boyunca genişletilmiştir. Söz grupları ve bazı noktalama işaretleri dikkate alınarak seçilen cümleler, ezgi özellikleri açısından incelenmiştir. Dört farklı formdaki (olumlu, olumsuz, olumlu soru ve olumsuz soru) toplam sekiz adet cümle ve noktalama işaretleri içeren dört adet cümle üzerinde uygulanan vurgu kuralları, sentezin doğallığını ve başarısını arttırmıştır.

Süre ayarı için; eğer sözcükteki difon sayısı 4'ten az ise, tüm difonların süresi %20 arttırılır. Bu işlemin amacı; anlaşılabilirliği arttırmaktır. Difon sayısı 6'dan fazla ise, tüm difonların süresi %5 azaltılır. Bu işlemin amacı da; uzun bir sözcüğün daha hızlı okutulmasıdır.

Cümle içinde yer alan söz grubu (öbek) vurgusu; en önemli vurgulardan biridir. Bu vurgu için; vurgulu difonların hem temel frekansları %30, hem de enerjileri %70 arttırılır. Öbek olmayan sözcük vurguları için, vurgulu hecede yer alan difonların sadece temel frekansları %10 arttırılır. Fiildeki vurgu için; fiilin vurgulu hecesindeki difonların temel frekansları %20, enerjileri %40 arttırılır. Ayrıca olumlu ve olumsuz cümleler, biten ezgi ile sonlandırılır. Bu amaçla; son heceyi oluşturan difonların temel frekansları ve enerjileri kademeli olarak (sırayla %10, %15, %20) azaltılır.

Difon tabanlı bir Türkçe metinden konuşma sentezleyici için, söz gruplarının, çekimli fiillerin ve noktalama işaretlerinin doğal konuşmadaki ezgi özelliklerini kazanması için şu adımlar önerilir:

- Olumlu fiil çekimlerinde ve sözcük vurgularında, yalnızca vurgulu difonların perde frekanslarının %10 artırılması,
- Olumsuz fiil çekimlerinde vurgulu difonların perde frekanslarının %20 artırılması,
- Olumlu soru formundaki fiil çekimlerinde, vurgunun yer aldığı difonların perde frekanslarının %20, enerjilerinin %40 artırılması,
- Olumsuz soru formundaki fiil çekimlerinde, vurgunun yer aldığı difonların perde frekanslarının %20 artırılması,
- Tamlamalarda, öbek vurgusunun yer aldığı difonların perde frekanslarının %30, enerjilerinin %70 artırılması,

- Noktalama işaretlerinde boşluk difonunun süresinin; virgül için %120, noktalı virgül için %240 ve iki nokta üst üste için %360 kadar artırılması önerilir.

Önerilen ezgi kuralları sentezin başarımını ortalama 1,45/5,00 puan artırmıştır. Belirlenen kurallar, vurgu yerleri belirli olan cümlelerin daha doğal sentezlenmesini sağlayacaktır.

## 6.1 TARTIŞMA

Bazı sentezlerde tüm difonlar denenmesine rağmen başarılı sonuç elde edilememesi, Türkçede başka difon tiplerinin de olduğunu bize düşündürmektedir.

Soru eki yazıda ayrı yazılırken, okunurken bitişik okunduğu için, sentezleme için bitişik yazılmalıdır.

Dinleme testleri az sayıda sentez üzerinden yapılmış görünse de, daha fazla sentez dosyası hazırlanıp rastgele seçilerek dinleyicilerin beğenisine sunulabilir. Her durumda dinleyicilerin dikkat ve sabırları için test dosyası sayısı sınırlı tutulmak zorundadır.

## 6.2 GELECEK ÇALIŞMALAR

Gelecek çalışma olarak; sentezleyicide difondan daha büyük konuşma parçalarının kullanılması önerilebilir.

Temel frekans değiştirmede diğer yöntemlerin başarısı incelenebilir. Vurguda etkili olan esas bileşen temel frekans olduğu için, bu parametrenin değiştirilmesi karşılaştırmalı olarak incelenebilir.

Bu alıřmada belirlenen ezgi kurallarının MBROLA veya Festival gibi hazır sentezleyicilere entegre edilip denenmesi nerilebilir. Bylece belirlenen vurgu kurallarının genel geerlięi sınanmıř olacaktır.

Fujisaki ezgi modelinin Trke iin oluřturulması, yeni bir tez konusu olabilir. Bu model, dilin gramer yapısını detaylı bir řekilde inceledięi iin, dilbilim uzmanlarıyla iřbirlięi yapılarak kapsamlı bir alıřma ortaya konabilir.

Grldę gibi, bu tez kapsamında yapılan alıřmalar, birkaç farklı doęrultuda geliřtirilmeye ve geniřletilmeye uygun grnmektedir.

## KAYNAKLAR

Anonim. 2012. Web siteleri:

<http://demo.reformo.net/baran/index.php> Erişim Tarihi: 08/02/2012

<http://demo.reformo.net/baran2/index.php> Erişim Tarihi: 08/02/2012

<http://demo.reformo.net/baran3/index.php> Erişim Tarihi: 08/02/2012

<http://demo.reformo.net/baran4/index.php> Erişim Tarihi: 08/02/2012

- Abdullahbeşe, E. 2001. Fundamental frequency contour synthesis for Turkish text-to-speech, Yüksek Lisans Tezi, Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü.
- Allen, J., Hunnicutt M.S. and Klatt, D. 1987. From text to speech: The MITalk system, Cambridge University Pres.
- Arısoy, E., Arslan, L. and Demiralp, M. 2004. Duration of Turkish vowels revisited, ICTL 2004.
- Ayhan, K. 1998. Text to speech synthesizer in Turkish using non parametric techniques, Yüksek Lisans Tezi, ODTÜ Fen Bilimleri Enstitüsü.
- Aydemir, T. ve Yılmaz A.E. 2010. Türkçe fiil çekimlerinde vurgu konumunu belirlemek için bir yazılım kütüphanesi, IEEE 18. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU 2010), s. 696-699, 22-24 Nisan 2010, Diyarbakır, Türkiye.
- Banguoğlu, T. 2007. Türkçenin Grameri, Türk Dil Kurumu Yayınları, Ankara.
- Başara, E. S. 1992. Yapay ses üretim yöntemleri, Yüksek Lisans Tezi, 53 s, KTÜ Fen Bilimleri Enstitüsü.
- Bozkurt, B. 2000. Reading aid for visually impaired (A Turkish text-to-speech system development), Yüksek Lisans Tezi, 67 s, Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü.
- Bulut, M., Busso, C., Yıldırım, S., Kazemzadeh, A., Lee, C. M., Lee, S. and Naratanan, S. 2005. Investigating the role of phoneme-level modifications in emotional speech resynthesis, Interspeech, pp. 801-804.
- Burkhardt, F., Audibert, N., Malatesta, L., Türk, O., Arslan, L. and Auberge, V. 2006. Emotional prosody – does culture make a difference?, Speech Prosody, Dresden-Germany, paper no. 207.
- Büyükaşkoğlu, K. 1992. Konuşma işaretlerinin analiz ve sentezi, Yüksek Lisans Tezi, 81 s, İTÜ Fen Bilimleri Enstitüsü.
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F. ve Vrecken O. 1996. The Mbrola project: towards a set of high quality speech synthesizers free of use for noncommercial purposes, Proceedings of ICSLP 96.
- Dutoit, T. 1997. An introduction to text-to-speech synthesis, Kluwer Academic Publishers, pp. 133-145.
- Dutoit, T. 1999. A short introduction to text-to-speech synthesis.
- Erer, M. S. 1994. Karma söz üretme yöntemi ile Türkçe yazılı metinden söze geçme, Yüksek Lisans Tezi, 113s, İTÜ Fen Bilimleri Enstitüsü.
- Ergenç, İ. 2002. Konuşma dili ve Türkçenin söyleyiş sözlüğü”, 486 s., Multilingual Yayınları.
- Eskidere, Ö. 2000. Yazılım tabanlı söz sentezleyici tasarımı, Yüksek Lisans Tezi, 72 s, Uludağ Üniversitesi Fen Bilimleri Enstitüsü.
- Fujisaki, H. 1983. Dynamic characteristics of voice fundamental frequency in speech

- and singing, part in the book: “The Production of Speech”, pp. 39-55.
- Gerçek, A. 1991. A TMS 5220 based speech synthesis development system, Yüksek Lisans Tezi, 110 s, ODTÜ Fen Bilimleri Enstitüsü.
- Goncharoff, V. and Gries, P. 1998. An algorithm for accurately marking pitch pulses in speech signals, IASTED International conference SIP ‘98, Nevada, USA.
- Gutierrez-Arriola, T.M., Montero, J.M., Saiz, D. and Pardo, J.M. 2001. New rule-based and data-driven strategy to incorporate Fujisaki’s F0 model to a text-to-speech system in Castilian Spanish, ICASSP 2001, pp. 821-824.
- Güner, E. ve Demiroğlu, C. 2011. Vurgu ve söyleyiş modelleme hatalarının SMM tabanlı Türkçe MKS sistemindeki etkileri, IEEE 19. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU-2011), s. 940-943, Kemer, Antalya.
- Güven, K. 1994. PC based speech synthesis for Turkish, Yüksek Lisans Tezi, 67s, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü.
- Inkelas, S. and Orgun, C.O. 2003. Turkish stress: a review, *Phonology*, 20, pp. 139–161.
- Kabak, B. and Vogel, I. 2005. Irregular stress in Turkish, Ms. University of Konstanz/ University of Delaware.
- Karlı, A. 2005. Örnek bir dizi cümle için Türkçe metinden konuşma sentezleme, Yüksek Lisans Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü.
- Külekçi, O. ve Oflazer, K. 2006. An infrastructure for Turkish prosody generation in text-to-speech synthesis”, 15th Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN 2006).
- Lemmetty, S. 1999. Review of speech synthesis technology, M.Sc. Thesis, Helsinki University of Technology, 104 p.
- McAuley R. J. ve Quatieri, T. F. 1986. Speech analysis/synthesis based on a sinusoidal representation, *IEEE Trans. Acoust. Speech Signal Process.*, 34, 744–754.
- Mixdorff, H. 1998. Intonation patterns of German – model-based quantitative analysis and synthesis of F0 contours”, Ph.D. thesis, TU Dresden.
- Moulines, M. ve Charpentier, F. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones”, *Speech Communication* volume: 9, pp. 453–467.
- Nabiyev, V. 2005. Yapay Zeka (Problemler – Yöntemler - Algoritmalar), Seçkin Yayınevi.
- Oskay, B., Salor, Ö., Özkan, Ö., Demirekler M. ve Çiloğlu T. 2001. Intonation abstraction from text and its applications for Turkish sentences (in Turkish), 9th IEEE Signal Processing and Communications Applications Symposium, SİU-2001, s. 238–243.
- Orhan, M. C. ve Demiroğlu, C. 2011. “Konuşmacı aradeğerlemeli SMM tabanlı metinden konuşma sentezleme sistemi”, IEEE 19. Sinyal İşleme ve İletişim Uygulamaları Konferansı (SİU-2011), s. 781-784, Antalya, Türkiye.
- Önür, Ç. 2001. Concatenative speech synthesis based on a sinusodal speech model, Yüksek Lisans Tezi, 50 s, ODTÜ Fen Bilimleri Enstitüsü.
- Özen, Ş. S. 2002. Türkçe konuşma sentezleme, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Elektrik ve Elektronik Mühendisliği Bölümü.
- Öztürk, Ö. 2005. Modelling phoneme durations and fundamental frequency contours in Turkish speech, Doktora Tezi, ODTÜ Fen Bilimleri Enstitüsü.

- Özüm, İ. Y. 1993. A speech synthesis system for Turkish language based on the concatenation of phonemes taken from a speaker, Yüksek Lisans Tezi, ODTÜ Fen Bilimleri Enstitüsü.
- Pasanen, A. 2001. Speech synthesis, seminer notları, 28 Ağustos 2001.
- Pertti, P. 2006. A review of articulatory speech synthesis, Master's Thesis, Helsinki University of Technology.
- Pierrehumbert, J. 1980. The phonology and phonetics of English intonation, Ph.D. Thesis, MIT.
- Rossi, P. S., Palmieri, F. and Cutugno F. 2002. A method for automatic extraction of Fujisaki-model parameters, *Proceedings of Speech Prosody*, pp. 615-618
- Salor, Ö. 1999. Signal processing aspects of text to speech synthesis in Turkish, Yüksek Lisans Tezi, ODTÜ Fen Bilimleri Enstitüsü.
- Salor, Ö., Pellom, B., Çiloğlu, T., Hacıoğlu, K. and Demirekler, M. 2002. On developing new text and audio corpora and speech recognition tools for the Turkish language, *ICSLP-2002: Inter. Conf. On Spoken Language Processing*, Denver, Colorado USA, pp. 349-352.
- Sezer, E. 1981. On non-final stress in Turkish, *Journal of Turkish Studies*-5, pp. 61-69.
- Sproat, R. 1998. Multilingual text-to-speech synthesis, *The Bell Labs Approach*, Kluwer Academic Publishers, Bölüm-6.
- Şayli, Ö. 2002. Duration analysis and modeling for Turkish text-to-speech synthesis, Yüksek Lisans Tezi, Boğaziçi Üniversitesi.
- Taylor, P. 1994. The rise-fall connection model of intonation, *Speech Communication* 15, pp. 169-186.
- 't Hart vd. 1990. A perceptual study of intonation – an experimental-phonetic approach to speech melody, Cambridge University Press.
- Uslu, İ.B. ve İlk, H.G. 2009. Türkçe metinden konuşma sentezlemede Fujisaki ezgi modeli, *IEEE 17. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU-2009)*, Antalya, s. 844-847.
- Uslu, İ.B., Yılmaz, A.E. ve İlk, H.G. 2011. Türkçe metinden konuşma sentezlemede fiil çekimleri için yeni bir ezgi modeli, *IEEE 19. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU-2011)*, Antalya, s. 638-641.
- Verhelst, W. 2000. Overlap-add methods for time-scaling of speech, *Speech Communication*, vol.30, s. 207-221, 2000.
- Vural, E. 2003. A prosodic Turkish text-to-speech synthesizer, Yüksek Lisans Tezi, Sabancı Üniversitesi, Fen Bilimleri Enstitüsü.
- Türk, O. 2000. New methods for voice conversion, Yüksek Lisans Tezi, Boğaziçi Üniversitesi.
- Yılmaz, Asım Egemen, 2009. Türkçe metinden konuşma sentezleme uygulamaları için bir veri sözlük seti ve yazılım çerçevesi önerisi, *IEEE 17. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU 2009)*, s. 956-959, Antalya, Türkiye.
- Zhou, G., Hansen, J.H.L ve Kaiser J.F. 2001. Nonlinear feature based classification of speech under stress, *IEEE Trans. On Speech and Audio Process.* Vol.9, no. 3.

## EK 1 Türkçe Metinlerin İstatistiksel Özellikleri

---

Türkçede bulunan 29 harfi aşağıdaki gibi sınıflandırırsak;

$V = \{a, e, ı, i, o, ö, u, ü\} : \text{ünlüler (Vowels)}$

$C = \{b, c, ç, d, f, g, ğ, h, j, k, l, m, n, p, r, s, ş, t, v, y, z\} : \text{ünsüzler (Consonants)}$

Ünsüzleri; sert ünsüzler ve yumuşak ünsüzler olarak iki gruba ayırabiliriz:

$C_{\text{soft}} = C_s = \{b, c, d, g, ğ, j, l, m, n, r, v, y, z\} : \text{yumuşak ünsüzler (C}_s\text{)} \text{ (ötümlü, Ergenç, 2002)}$

$C_{\text{hard}} = C_h = \{ç, f, h, k, p, s, ş, t\} : \text{sert ünsüzler (C}_h\text{)} \text{ (ötümsüz, Ergenç, 2002)}$

$$C = C_s \cup C_h$$

Ayrıca, ünsüzleri sürekli ve süreksiz ünsüzler olarak da iki gruba ayırabiliriz:

$C^{\text{sustainable}} = C^s = \{f, ğ, h, j, l, m, n, r, s, ş, v, y, z\} : \text{sürekli ünsüzler (C}^s\text{)}$

$C^{\text{unsustainable}} = C^u = \{b, c, ç, d, g, k, p, t\} : \text{süreksiz ünsüzler (C}^u\text{)}$

$$C = C^s \cup C^u$$

$C_h^s = C_h \cap C^s = \{f, h, s, ş\} : \text{sert sürekli ünsüzler (C}_h^s\text{)}$

$C_h^u = C_h \cap C^u = \{ç, k, p, t\} : \text{sert süreksiz ünsüzler (C}_h^u\text{)}$

$C_s^s = C_s \cap C^s = \{ğ, j, l, m, n, r, v, y, z\} : \text{yumuşak sürekli ünsüzler (C}_s^s\text{)}$

$C_s^u = C_s \cap C^u = \{b, c, d, g\} : \text{yumuşak süreksiz ünsüzler (C}_s^u\text{)}$

Böylece, 4 adet ünsüz alt grubu ve 1 adet ünlü grubu elde edilmiş olur.

$v \in V, c \in C$  olmak üzere;

bu grupların tüm kombinasyonları ve bu kombinasyonlara örnekler çizelge-1'de verilmektedir:

Çizelge 1 Digram kombinasyonları ve örnekler

|                    |                           |
|--------------------|---------------------------|
| $x(c_s^s)(c_h^s)x$ | : emsal, yansı, Tevfik    |
| $x(c_s^s)(c_h^u)x$ | : imkan, çarşı, şarkı     |
| $x(c_s^s)(c_s^s)x$ | : anla, dövme, sazlı      |
| $x(c_s^s)(c_s^u)x$ | : amca, evcil, silgi      |
| $x(c_s^s)(v)x$     | : makarna, vize, yuva     |
| $x(c_s^u)(c_h^s)x$ | : adsız, yadsıma          |
| $x(c_s^u)(c_h^u)x$ | : -                       |
| $x(c_s^u)(c_s^s)x$ | : abla, kadrın, podyum    |
| $x(c_s^u)(c_s^u)x$ | : ecdad, teşebbüs, tedbir |
| $x(c_s^u)(v)x$     | : sancı, damak, bugün     |
| $x(c_h^s)(c_h^s)x$ | : müessese, ishal, yassı  |
| $x(c_h^s)(c_h^u)x$ | : aşçı, askı, hastane     |
| $x(c_h^s)(c_s^s)x$ | : ihmal, tasma, safra     |
| $x(c_h^s)(c_s^u)x$ | : meşgul, sohbet, teşbih  |
| $x(c_h^s)(v)x$     | : hatıra, şölen, ahır     |
| $x(c_h^u)(c_h^s)x$ | : eksen, tepsi, tekfen    |
| $x(c_h^u)(c_h^u)x$ | : teşekkür, iptal, etki   |
| $x(c_h^u)(c_s^s)x$ | : açma, toplu, toprak     |
| $x(c_h^u)(c_s^u)x$ | : ikbal                   |
| $x(c_h^u)(v)x$     | : kitapçı, tutku, peçete  |
| $x(v)(c_h^s)x$     | : esir, aşık, pahalı      |
| $x(v)(c_h^u)x$     | : patates, seçkin, ukala  |
| $x(v)(c_s^s)x$     | : kol, ozan, düğme        |
| $x(v)(c_s^u)x$     | : dede, abla              |
| $x(v)(v)x$         | : saat, aile, müessese    |

Bu noktada digram ve trigram tanımlarını verelim. 29 harfin ikişerli kombinasyonlarına “digram”; üçerli kombinasyonlarına da “trigram” adı verilir.

Bu bölümde, 3 adet Türkçe elektronik kitaptan (E-kitap) elde edilen istatistikler ışığında, süre modelleme çalışmamız için bir yol haritası oluşturulmaya çalışılacaktır. İncelenen kitapların özellikleri çizelge 2’de verilmektedir.

Çizelge 2 İncelenen Türkçe E-kitapların <sup>(1)</sup> özellikleri

|                 | <b>Kitabın Adı / Yazarı / Türü</b>                           | <b>Sayfa sayısı</b> | <b>Sözcük sayısı</b> |
|-----------------|--------------------------------------------------------------|---------------------|----------------------|
| <b>E-kitap1</b> | Veda Yemeği / Michel Tournier / Çeviri                       | 111                 | 52537                |
| <b>E-kitap2</b> | Son Antlaşma / Can Eryümlü / Roman                           | 250                 | 95988                |
| <b>E-kitap3</b> | Bir Hasta Sahibinin Hastane Günlüğü / Doğan Pazarcıklı / Anı | 40                  | 16472                |

(1): Kitaplara [www.altkitap.com](http://www.altkitap.com) adresinden ulaşılmıştır.

E-kitaplar digram ve trigram dağılımları açısından incelenmiş, elde edilen sonuçlar çizelge 4 ve çizelge 5 ile verilmiştir. Görülmektedir ki; dilimizde ( $\mathbf{v-c_s^s}$ ), ( $\mathbf{c_s^s-v}$ ) ve ( $\mathbf{c_s^u-v}$ ) gruplarında yer alan digramlar, tüm digramların % 60'tan biraz fazlasını oluşturmaktadır. Benzer şekilde, ( $\mathbf{v-c_s^s-v}$ ), ( $\mathbf{c_s^s-v-c_s^s}$ ) ve ( $\mathbf{v-c_s^u-v}$ ) gruplarında yer alan trigramlar da tüm trigramların % 40'a yakınıni oluşturuyor. Bundan dolayı, yapılacak olan modelleme çalışmalarına bu gruplarla başlamanın avantajlı olacağı düşünülmektedir. Eğer bu gruplar üzerinde başarılı sonuçlar elde edilirse, bu başarı geniş bir küme için elde edilmiş olacaktır. Dolayısıyla, hem erkek, hem de kadın sesi ile kaydedilen difonlar, öncelikle bu gruplardan seçilmiştir.

Çizelge 3 Digram gruplarının yüzdeleri

| <b>Grup</b>                 | <b>E-kitap1</b> | <b>E-kitap2</b> | <b>E-kitap3</b> |
|-----------------------------|-----------------|-----------------|-----------------|
| $x(\mathbf{v-c_s^s})_x$     | 28,15%          | <b>28,40 %</b>  | 28,72 %         |
| $x(\mathbf{c_s^s-v})_x$     | 21,66%          | <b>21,95 %</b>  | 21,60 %         |
| $x(\mathbf{c_s^u-v})_x$     | 11,15%          | <b>11,54 %</b>  | 11,18 %         |
| $x(\mathbf{c_h^u-v})_x$     | 7,38 %          | <b>6,94 %</b>   | 7,29 %          |
| $x(\mathbf{v-c_h^u})_x$     | 5,67%           | <b>5,35 %</b>   | 5,73 %          |
| $x(\mathbf{c_h^s-v})_x$     | 5,15 %          | <b>5,10 %</b>   | 5,90 %          |
| $x(\mathbf{c_s^s-c_s^s})_x$ | 3,84 %          | <b>4,17 %</b>   | 3,52 %          |
| $x(\mathbf{v-c_h^s})_x$     | 3,96%           | <b>4,16 %</b>   | 4,33 %          |

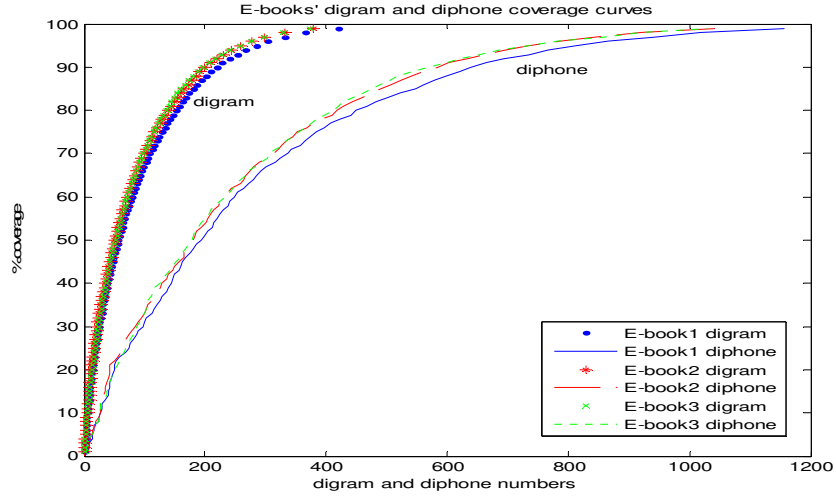
Çizelge 3 Digram gruplarının yüzdeleri (devam)

|                                        |        |               |        |
|----------------------------------------|--------|---------------|--------|
| $x(\mathbf{c}_s^s - \mathbf{c}_s^u)_x$ | 3,38 % | <b>3,11 %</b> | 2,82 % |
| $x(\mathbf{v} - \mathbf{c}_s^u)_x$     | 2,38%  | <b>2,77 %</b> | 2,71 % |
| $x(\mathbf{c}_h^u - \mathbf{c}_s^s)_x$ | 1,34 % | <b>1,29 %</b> | 1,16 % |
| $x(\mathbf{c}_s^s - \mathbf{c}_h^u)_x$ | 1,02 % | <b>0,98 %</b> | 0,91 % |
| $x(\mathbf{c}_h^s - \mathbf{c}_h^u)_x$ | 0,84 % | <b>0,90 %</b> | 1,22 % |
| $x(\mathbf{c}_h^s - \mathbf{c}_s^s)_x$ | 0,54 % | <b>0,82 %</b> | 0,47 % |
| $x(\mathbf{c}_h^u - \mathbf{c}_h^u)_x$ | 0,84 % | <b>0,74 %</b> | 0,82 % |
| $x(\mathbf{c}_s^s - \mathbf{c}_h^s)_x$ | 0,58 % | <b>0,67 %</b> | 0,72%  |
| $x(\mathbf{v} - \mathbf{v})_x$         | 1,32%  | <b>0,42 %</b> | 0,39 % |
| $x(\mathbf{c}_h^s - \mathbf{c}_h^s)_x$ | 0,19 % | <b>0,21 %</b> | 0,10 % |
| $x(\mathbf{c}_s^u - \mathbf{c}_s^s)_x$ | 0,16%  | <b>0,17 %</b> | 0,10 % |
| $x(\mathbf{c}_h^u - \mathbf{c}_h^s)_x$ | 0,22 % | <b>0,15 %</b> | 0,20 % |
| $x(\mathbf{c}_s^u - \mathbf{c}_s^u)_x$ | 0,10%  | <b>0,06 %</b> | 0,05 % |
| $x(\mathbf{c}_h^u - \mathbf{c}_s^u)_x$ | 0,07 % | <b>0,04 %</b> | 0,02 % |
| $x(\mathbf{c}_s^u - \mathbf{c}_h^s)_x$ | 0,04%  | <b>0,03 %</b> | 0,01 % |
| $x(\mathbf{c}_s^u - \mathbf{c}_h^u)_x$ | 0,01%  | <b>0,02 %</b> | 0,00 % |
| $x(\mathbf{c}_h^s - \mathbf{c}_s^u)_x$ | 0,03 % | <b>0,02 %</b> | 0,02 % |

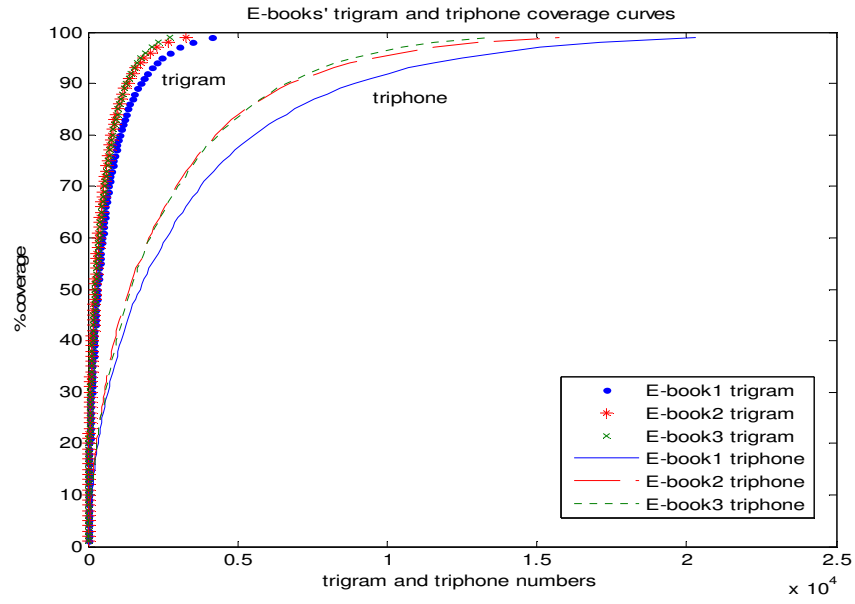
Çizelge 4 En çok rastlanan trigram gruplarının yüzdeleri

| Grup                                                | E-kitap1 | E-kitap2 | E-kitap3 |
|-----------------------------------------------------|----------|----------|----------|
| $x(\mathbf{v} - \mathbf{c}_s^s - \mathbf{v})_x$     | % 16.30  | % 16.92  | % 16.97  |
| $x(\mathbf{c}_s^s - \mathbf{v} - \mathbf{c}_s^s)_x$ | % 13.77  | % 14.02  | % 14.37  |
| $x(\mathbf{v} - \mathbf{c}_s^u - \mathbf{v})_x$     | % 7.90   | % 7.87   | % 8.25   |

Digram-difon ve trigram-trifon kapsama grafikleri şekil 1 ve şekil 2’de verilmiştir.



Şekil 1 E-kitap’ların digram ve difon kapsama grafiği



Şekil 2 E-kitap’ların trigram ve trifon kapsama grafiği

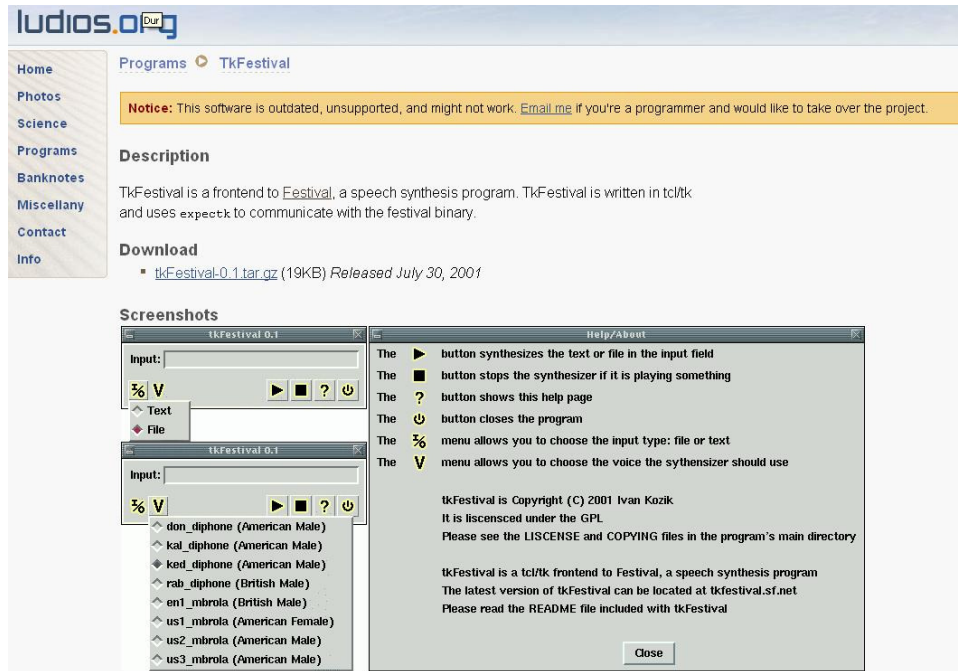
## EK 2 Diğer Dillerde Yapılan MKS Çalışmaları

Bu bölümde diğer dillerde geliştirilen bazı MKS sistemleri tanıtılmaktadır. Sistemlerin özellikleri tablo ile özetlenmeye çalışılmış, ayrıca şekilleri de verilmiştir.

- **Festival TTS / TkFestival**

Çizelge 1 Festival MKS sisteminin özellikleri

| Özellik              | İçerik                                   |
|----------------------|------------------------------------------|
| Desteklediği Diller: | İngilizce, İspanyolca                    |
| Türü                 | Ticari                                   |
| Sentezleme Yöntemi:  | Difon birleştirme (Residual Excited LPC) |
| Yazıldığı Dil:       | C++                                      |



Şekil 1 TkFestival sistemi

- **AT&T Labs Research**

Çizelge 2 AT&T Labs MKS sisteminin özellikleri

| Özellik              | İçerik                                    |
|----------------------|-------------------------------------------|
| Desteklediği Diller: | İngilizce, İspanyolca, Fransızca, Almanca |
| Türü                 | Araştırma / Ticari                        |
| Sentezleme Yöntemi:  | Eklemeli ve formant sentezleme            |
| Yazıldığı Dil:       | Java                                      |

**AT&T Labs Text-to-Speech: Demo - Microsoft Internet Explorer**

Dosya Düzen Görünüm Sık Kullanılanlar Araçlar Yardım

Geri - - - - - Ara Sık Kullanılanlar - - - - -

Adres <http://www.research.att.com/~ttsweb/tts/demo.php>

**at&t** | AT&T Labs, Inc. - Research

**Text-To-Speech (TTS) -- Our Demo Speaks Your Text**

Home | > Demo | FAQ | Publications | Contact  
[Wizzard Software](#) | [AT&T Natural Voices](#)  
[TV Commercial](#)

**STEP 1 Voice & Language:** Crystal ... US English

**STEP 2 Text:** [ [Selected language only](#) | [300 character limit](#) | [Help with UTF-8 or Latin-1](#) ]

**STEP 3 Click:** SPEAK - or - DOWNLOAD [ [restrictions apply](#)\* ]

Something mispronounced? SEND FEEDBACK

Something broken? Please consult the FAQ page [Diagnostics](#) before sending details to [tts-feedback](#).

Comments, suggestions and questions not answered in the [FAQ](#) should also be sent to [tts-feedback](#).

\* **This page is for demonstration purposes only.**  
 Audio samples are for private, non-commercial use. Server resources are limited and are reserved for this website; direct use of the CGI scripts is not allowed. See the [usage policy](#) for details.

Şekil 2 AT&T Labs sistemi

- IBM TTS

Çizelge 3 IBM MKS sisteminin özellikleri

| Özellik              | İçerik                           |
|----------------------|----------------------------------|
| Desteklediği Diller: | İngilizce, Portekizce, Fransızca |
| Türü                 | Ticari                           |
| Sentezleme Yöntemi:  | Eklemeli ve formant sentezleme   |
| Yazıldığı Dil:       | Java                             |

Adres <http://www-306.ibm.com/software/pervasive/tech/demos/tts.shtml>

Search the Web  Google Yahoo! Ask.com LookSmart Customize this toolbar Highlight

IBM Country/region [select] Terms of use Search

Home Products Services & solutions Support & downloads My account

Software > Wireless - Voice - Pervasive > WebSphere Voice >

## WebSphere Voice

WebSphere software

### Technology

Welcome to the IBM Text-To-Speech Demonstration.

**This demonstration is being performed using 8-bit u-Law at 8KHz which is what one would hear using a telephone.**

IBM's speech synthesis technology, ViaVoice Text-To-Speech, is currently available in many products and platforms:

- ViaVoice desktop software, providing speech synthesis capability for applications on Windows, Unix, and Linux platforms.
- IBM Embedded ViaVoice Text-To-Speech, which provides speech synthesis for mobile devices.
- IBM Text-To-Speech is also a component of WebSphere Voice Server, IBM's speech recognition and synthesis system for telephony networks.

**Please enter the text in the language that you want spoken. The demo will not translate your text.**

**Text to speak (Limited to 128 characters).**

Enter your text in the desired language and then press 'Speak'

Speak Clear

Language US English Male

Volume (0-100) 92

Şekil 3 IBM MKS sistemi

- **SourceForge.net eSpeak**

Çizelge 4 eSpeak MKS sisteminin özellikleri

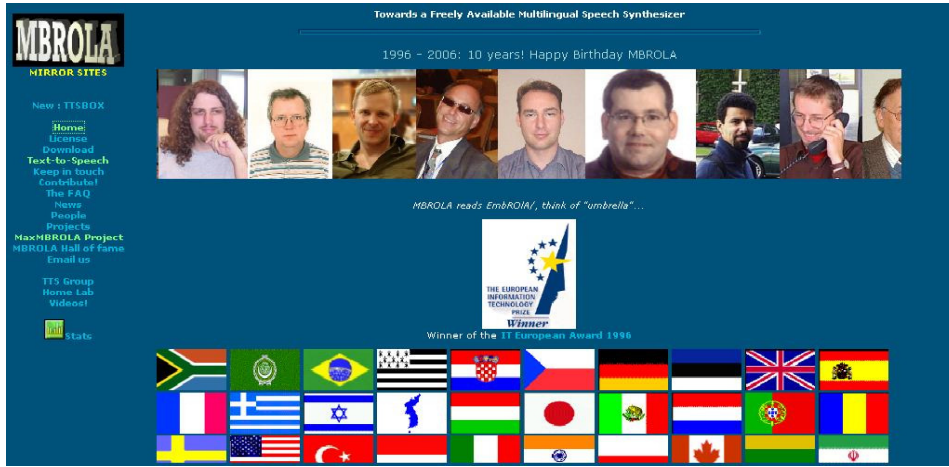
| Özellik              | İçerik                                                               |
|----------------------|----------------------------------------------------------------------|
| Desteklediği Diller: | İngilizce, Almanca, Esperanto (halklar arası evrensel iletişim dili) |
| Türü                 | Ticari                                                               |
| Sentezleme Yöntemi:  | Eklemeli                                                             |
| Yazıldığı Dil:       | C++                                                                  |

- **GVZ Ses Teknolojileri**



Şekil 4 GVZ MKS sistemi

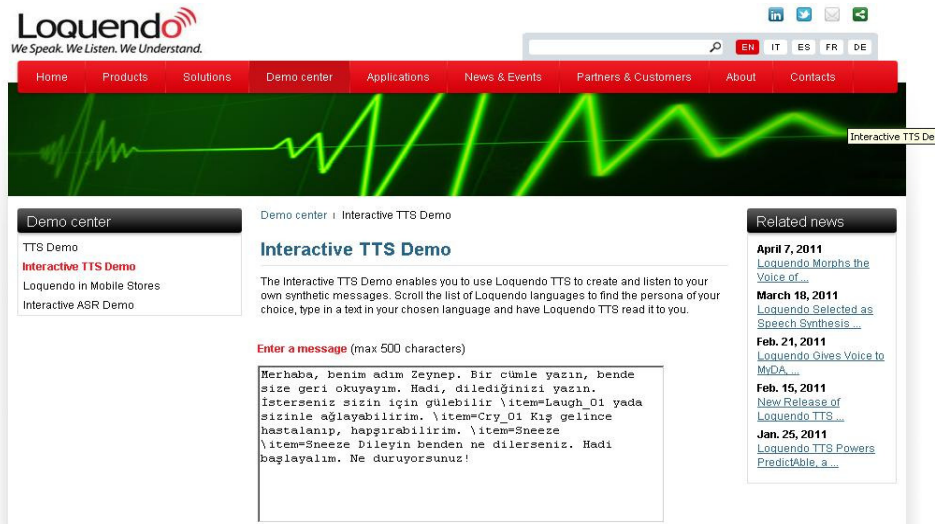
- **MBROLA Multilingual Speech Synthesis**



Şekil 5 MBROLA, Multilingual Speech Synthesizer Project

- **Loquendo TTS**

Loquendo TTS; Türkçe dâhil pek çok dilde MKS özelliği bulunan ve sentezin içine gülme, ağlama, esneme gibi duygusal etkiler katabilen başarılı bir program. Programın ağ sayfası üzerinden sentez yapabilmek de mümkün.



Şekil 6 Loquendo konuşma sentezleyici

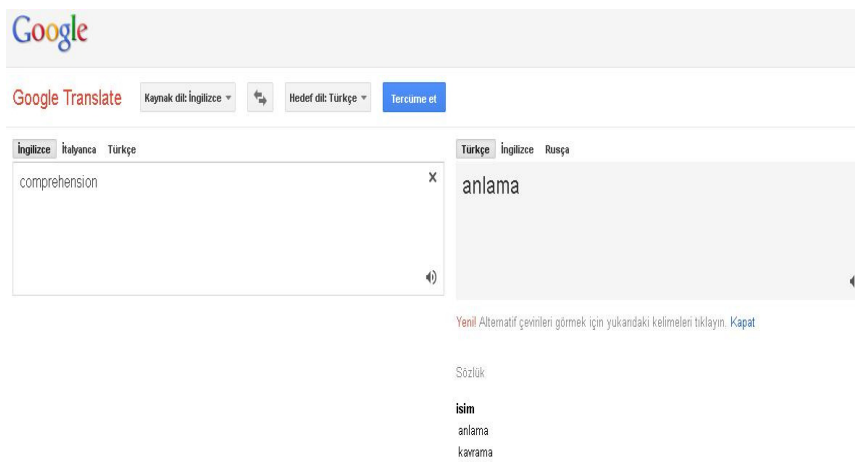
- **Acapela TTS**

Türkçe dâhil pek çok dilde MKS özelliği olan bu sistemin konuşma kalitesi oldukça iyi. Noktalama işaretlerinin etkisi de güzel düşünülmüş.



Şekil 7 Acapela konuşma sentezleyici

- **Google Translate**



Şekil 8 Google Translate’de kullanılan konuşma sentezleyici

---

Çizelge 5 Bazı Ticari MKS Ürünleri

|   |                                                                                                                                                                                                               |
|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Acapela TTS<br><a href="http://www.acapela-group.com/text-speech-contact.html">http://www.acapela-group.com/text-speech-contact.html</a> (son erişim: 05.03.2012)                                             |
| 2 | Balabolka TTS<br><a href="http://www.cross-plus-a.com/balabolka.htm">http://www.cross-plus-a.com/balabolka.htm</a> (son erişim: 05.03.2012)                                                                   |
| 3 | Braille Teknik Ltd. Şti.<br><a href="http://www.brailleteknik.com/jaws.html">http://www.brailleteknik.com/jaws.html</a> (son erişim: 05.03.2012)                                                              |
| 4 | DİKTE Yöndata Bilgisayar Ltd. Şti.<br><a href="http://www.dikte.com.tr/metinokuma.php">http://www.dikte.com.tr/metinokuma.php</a> (son erişim: 05.03.2012)                                                    |
| 5 | Google translate<br><a href="http://translate.google.com">http://translate.google.com</a> (son erişim: 05.03.2012)                                                                                            |
| 6 | <a href="http://www.gvz.com.tr/index.html">GVZ Ses tanıma ve sentezleme teknolojileri şirketi</a><br><a href="http://www.gvz.com.tr/index.html">http://www.gvz.com.tr/index.html</a> (son erişim: 05.03.2012) |
| 7 | IBM WebSphere Voice<br><a href="http://www-01.ibm.com/software/pervasive/tech/demos/tts.shtml">http://www-01.ibm.com/software/pervasive/tech/demos/tts.shtml</a> (son erişim: 05.03.2012)                     |
| 8 | Loquendo S.p.A., a Telecom Italia Group Company<br><a href="http://www.loquendo.com/en/demo-center/tts-demo/">http://www.loquendo.com/en/demo-center/tts-demo/</a> (son erişim: 05.03.2012)                   |

## ÖZGEÇMİŞ

**Adı Soyadı:** İbrahim Baran Uslu  
**Doğum Yeri:** Ankara  
**Doğum Tarihi:** 18.02.1976  
**Medeni Hâli:** Bekâr  
**Yabancı Dili:** İngilizce

### Eğitim Durumu (Kurum ve Yıl)

**Lise:** Özel Arı Fen Lisesi (1991-1994)

**Lisans:** Ankara Üniversitesi Elektronik Mühendisliği Bölümü (1994-1998)

**Yüksek Lisans:** Başkent Üniversitesi Elektrik-Elektronik Mühendisliği Bölümü (2000-2003)

**Doktora:** Ankara Üniversitesi Elektronik Mühendisliği (2003-2012)

### Çalıştığı Kurum/Kurumlar ve Yıl

Arş. Gör., Hacettepe Üniversitesi Elektrik ve Elektronik Mühendisliği Bölümü, 1999-2000.

Arş. Gör., Başkent Üniversitesi Elektrik-Elektronik Mühendisliği Bölümü, 2000-2006.

Öğr. Gör., Başkent Üniversitesi Elektrik-Elektronik Mühendisliği Bölümü, 2006-...

### Yayınları

Uslu, İ. B. ve İlk, H. G., Türkçe metinden konuşma sentezlemede kural tabanlı duyumsal bir ezgi modeli, IEEE 20. Sinyal İşleme ve İletişim Uygulamaları Sempozyumu SİU-2012, 18-20 Nisan 2012, Fethiye-Antalya (sözlü sunum).

Uslu, İ. B., Demir, N. İlk, G ve Yılmaz, A. E., Bilgisayar bir metni vurgulu okuyabilir mi?", BİLİG Türk Dünyası Sosyal Bilimler Dergisi, (63-Güz, Eylül 2012 sayısında yayımlanmak üzere kabul edildi).

Uslu, İ. B., İlk, H. G. Ve Yılmaz, A. E., Türkçe metinden konuşma sentezlemede doğallığın artırılması için öneriler", EMO Bilimsel Dergi, cilt: 1, sayı: 2, s. 95-102, Aralık 2011.

Uslu, İ. B., Yılmaz, A. E. ve İlk, H. G., Türkçe metinden konuşma sentezlemede fiil çekimleri için yeni bir ezgi modeli”, IEEE 19. Sinyal İşleme ve İletişim Uygulamaları Sempozyumu SİU-2011, s. 638-641, 20-22 Nisan 2011, Kemer-Antalya (sözlü sunum).

Uslu, İ. B., Yılmaz, A. E. and İlk, H. G., Statistical digram and trigram analysis of Turkish in terms of coverage and entropy for possible language and speech based applications, EUSIPCO 2010, s. 776-780, 22-26 Ağustos 2010, Aalborg-Danimarka (poster bildirisi).

Uslu, İ. B. ve İlk, H. G., Türkçe metinden konuşma sentezlemede Fujisaki ezgi modeli”, IEEE 17. Sinyal İşleme ve Uygulamaları Kurultayı (SİU 2009), s. 844–847, 9-11 Nisan 2009, Antalya (sözlü sunum).