5.2. HESAPLAMALI SESBİLGİSİNDE EZGİ SENTEZLEME

İpek Pınar UZUN
Ankara Üniversitesi, Dilbilim Bölümü
pinarbekar@gmail.com

1. Hesaplamalı Sesbilgisi Alanına Genel Bakış

Konuşma seslerinin fiziksel ve anatomik özelliklerinin tanımlanması ve konuşma sırasında ses dalgalarındaki temel değişimlerin betimlenmesini ve algılanmasını inceleyen sesbilgisi (*phonetics*) alanı, akustik sesbilgisi (*acoustic phonetics*), işitsel sesbilgisi (*auditory phonetics*) ve eklemleyici sesbilgisi (*articulatory phonetics*) olmak üzere üç temel alt alana ayrılmaktadır. Bu yazının ana konularından birini oluşturan akustik sesbilgisi, konuşma seslerinin hava basıncına dayalı değişimlerini, fizik ve matematik gibi disiplinlerle birlikte incelemektedir. Son yıllarda akustik sesbilgisi alanında yapılmakta olan araştırmalar, mühendislik tabanlı istatistiksel çözümleme teknikleriyle birlikte sesin fiziksel özelliklerinin belirlenmesi ve sesbilimsel sorunlara bu tekniklerle yanıt aranması gerekliliğini ortaya koymuştur. Doğal dil işleme sürecine bilgisayar kullanımı ile istatistiksel bir çözümleme getirmiş olan hesaplamalı dilbilim (*computational linguistics*) alanı ile akustik sesbilgisi alanlarındaki çözümleme tekniklerinin biraraya getirilmesiyle, yeni bir disiplinlerarası alt alanının oluşumuna zemin hazırlanmıştır. Hesaplamalı sesbilgisi (*computational phonetics*) olarak adlandırılan bu disiplinlerarası alan, fiziksel bir birim olan sesin perde değerinin, enerji yoğunluğunun, süresinin ve formant sıklıklarının belirlenmesi gibi akustik sesbilgisel sorunları, algoritma tabanlı çözümleme teknikleri geliştirerek, matematiksel gözlem yöntemiyle inceleyen bir araştırma alanıdır.

Algoritma temelli dizilere dayalı çözümleme teknikleri kullanılarak, konuşma seslerinin sentezlenmesi (*speech synthesis*), konuşmacı tanıma/tanımlama (*speech recognition*), metinden konuşma sentezleme (*text-to-speech synthesis*) gibi pek çok alanın da tabanını oluşturan hesaplamalı sesbilgisine ilişkin analiz teknikleri, günümüzde ham veri ve istatistiksel veri analizi biçiminde ele alınmaktadır. Boersma ve Weenink (2020) tarafından sesbilim alanyazınına kazandırılan ve son yıllarda önemli bir araştırmacı kitlesi kazanan Praat ses analizi yazılımındaki komut dizileri

kullanılarak, günümüzde ham veri analizleri gerçekleştirilebilmektedir. İşlenmiş verilere ilişkin sorunlara, hesaplamalı sesbilgisindeki çeşitli istatistiksel algoritmalar temelinde geliştirilmiş analizlerle yanıt aranmaktadır. Bilgisayarbilimlerine (*computer sciences*) yönelik analiz tekniklerinin giderek önem kazandığı günümüz çağında, sesin bilgisayara tanımlanması konusunda hesaplamalı sesbilgisi alanının önemli bir rolü bulunmaktadır. Algoritma tabanlı bu analizlerdeki temel amaç, insana ait konuşma sesleri ile makine tarafından üretilmiş ait konuşma sesleri arasındaki farklılıkları en az indirgeyerek, makine sesini mümkün olduğunca en doğal biçime dönüştürmektir.

5. Hesaplamalı Sesbilgisi bölümünün ikinci kısmını oluşturan bu bölümde, sesbilim alanında önemli bir yeri olan ezgi sentezleme üzerine geçmişten günümüze kadar geliştirilmiş olan hesaplamalı ve akustik sesbilgisi tabanlı kuramsal yaklaşımlar ve modellerin alanyazın tanıtımı yapılacaktır. Bölümün ilk kısımlarında ezgi ve temel bileşenlerinin işleyişi üzerinden genel yaklaşımlar sunulacak, izleyen kısımlarında ise öncü ve günümüz kuramsal yaklaşımlarında ezgi sentezlemenin rolleri üzerinde durulacaktır. Bölümün son kısımlarında ise, bilgisayar-makine arakesinde metinden konuşma sentezleme gibi ezgiyi araç olarak kullanarak alanlar ele alınarak, günümüze kadar Türkçe üzerine yapılmış çalışmalardan söz edilmesi planlanmıştır.

2. Ezgi ve Temel Akustik Bileşenleri

Beynin sağ yarıküresinde konumlandığı varsayılan bürün (*prosody*) dizgesinin alt bir bileşeni olan ezgi (*intonation*) (bkz. Friederici & Alter, 2004), odak (*focus*), ton (*tone*), ses rengi (*timbre*) ve durak (*pause*) gibi alt bileşenlerden oluşmaktadır. Akustik sesbilgisel çözümleme teknikleriyle incelenebilen bu alt bileşenlerinin toplamı, Crystal (1992)'de de belirtildiği gibi, ilgili tümcenin ezgisini oluşturmaktadır. Ses tonundaki değişimlere dayalı olarak yükselme ve alçalma gösteren ezginin belirlenebilmesiyle, tümcedeki anlamsal ya da sözdizimsel belirsizlikler ortadan kaldırılabilmektedir. Söylem (*discourse*) alanında da inceleme konusu oluşturan ezgi, konuşucunun ses rengine dayalı değişimleri duyguduruma ya da ilgili sözcenin içinde bulunduğu bağlama dayalı olarak ele almaktadır. Yazının bir sonraki bölümünde ezginin temel bileşenlerinin akustik sesbilgisi ve sesbilim alanlarındaki görünümleri ve inceleme yöntemleri üzerinde durulması planlanmaktadır.

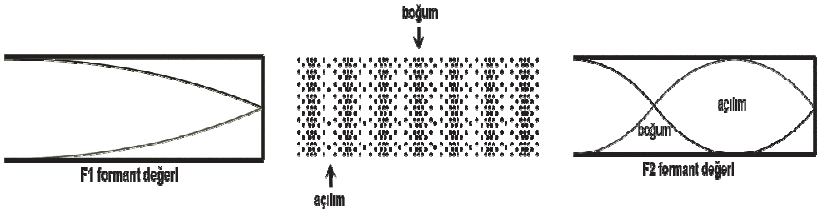
2.1. Temel Sıklık (F_0) ve Perde Vurguları (Hz)

Temel sıklık/F_0 (*fundamental frequency/F_0*) ve perde vurgusu (*pitch accent*) değerleri, Hertz (Hz) ile ölçülebilir nitelikteki akustik değiştirgenlerdir. Ses tellerindeki titreşim sıklığı, kısalık uzunluk gibi çeşitli fiziksel değerlere dayalı olarak belirlenen perde değeri, 0-500 Hz arasında cinsiyet farklığına dayalı belirlenebilmektedir.

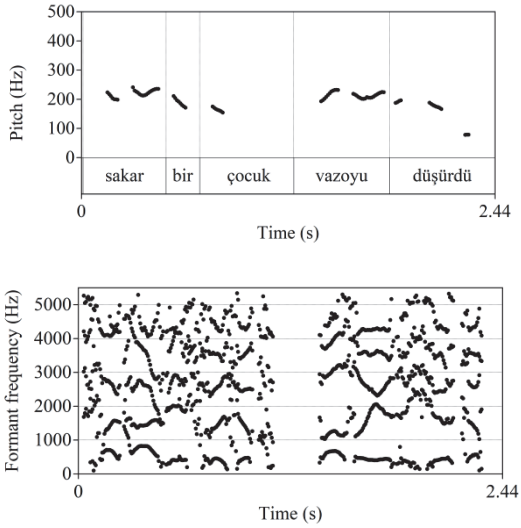
Yaklaşık değerleri açısından kadınlarda 170-210 Hz, erkeklerde 120-140 Hz ve çocuklarda 220-280 Hz olarak belirlenen F_0 değeri perdedeki en düşük sıklık değerini sunmaktadır (bkz. Kent & Read, 2002). Bu bölümdeki temel akustik değiştirgenlerden biri olan F_0 ve perde vurgusu değerleri, sözcükteki seslemsel birimler arasındaki akustik farklılıkların belirlenmesinde etkin rol oynamaktadır.

2.2. F_1 ve F_2 Formant Değerleri (Hz)

Birinci formant F_1 ve ikinci formant F_2 değerleri, Kent ve Read (2002)'de de belirtildiği gibi, Hertz (Hz) ile ölçülebilir nitelikli, konuşma esnasında ses yolunda ortaya çıkan değişimlerde ünlülerin fiziksel değerlerini belirleme amacıyla kullanılan bir akustik değiştirgendir. Bu bölümde, formant değerleri Chiba ve Kajiyama (1941)'de geliştirilmiş olan, ünlülerin akustik özelliklerinin modelleme amacıyla hava basıncı ve hız arasındaki bağıntıyı inceleyen Daralma Kuramı (*Perturbation Theory*) çerçevesinde ele alınmaktadır. Buna göre, hava moleküllerinin hareket etmedikleri alanlarda ses dalgaları boğum (*node*), yoğun oldukları alanlarda ise hareket fazla olduğu için açılım (*antinode*) olarak tanımlanmaktadır (Kent & Read, 2002; Malkoç, 2011; Ergenç & Bekar Uzun, 2017 gibi) (bkz. Şekil 1). Bu akustik kuram çerçevesinde, boğum alanları ünlülerinin çıkarılma özelliklerinin (boğumlanma) belirlendiği alanlar olarak tanımlanabilmektedir. Formant değerleri, Daralma Kuramında bir tüp modellemesi bağlamında incelenmekte ve hava moleküllerinin yer değiştirilme özelliklerine göre frekans farklılıkları ortaya konmaktadır.



Şekil 1. F_1 ve F_2 formant frekansları ve hava moleküllerinin hareketleri

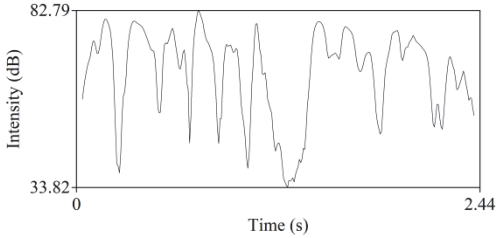


Şekil 2. "Sakar bir çocuk vazoyu düşürdü" tümcesinin ezgi görünümü (üstte) ve hava moleküllerinin hareketleri (altta)

Yukarıdaki şekillerde de gözlemlenebileceği gibi (bkz. Şekil 2), hava moleküllerindeki daralma sonucu, frekansın boğumlanma alanı genişlemekte ve frekans değeri yükselmektedir. Buna göre, ünlülerin belirlenmesinde etkili olan F_1 formantı 500-1000 Hz aralığında, F_2 formantı 1000-2000 Hz aralığında değişim gösterirken; kalan F_3, F_4, F_5 formantları 1000'in katları biçiminde artmaktadır.

2.3. Enerji Yoğunluğu (dB)

Enerji yoğunluğu (*intensity*) değeri, ses dalgasındaki genliğin artmasıyla doğru orantılı bir işleyişe sahip olan akustik bir değiştirgendir. Desibel (dB) ile ölçülebilir nitelikteki yoğunluk değeri, aynı oranda titreyen iki ses dalgasında genlik oranı ne kadar fazla ise, o ses dalgasının yoğunluk değeri de bir o kadar artmaktadır. Enerji yoğunluğu değerinin sağlıklı bireylerde genellikle 70-80 dB aralığında olması beklenmektedir (Kent & Read, 2002; Ergenç & Bekar Uzun, 2017).

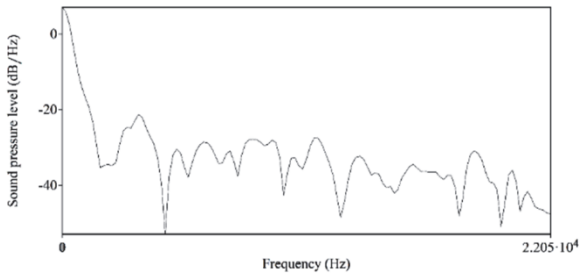


Şekil 3. "Sakar bir çocuk vazoyu düşürdü" tümcesinin enerji yoğunluğu

Şekil (3)'teki "*Sakar bir çocuk vazoyu düşürdü*" tümcesine ait enerji yoğunluğu grafiğinde de görüldüğü gibi, özellikle sözcük vurgusunun soluk baskısının yoğunlaşma alanlarının betimlenmesi ve alınan ses verisinin uygun bir ses kayıt ortamında alınıp alınmadığının belirlenmesi gibi akustik süreçlerde yoğunluk önemli bir ayırıcı işlev taşımaktadır.

2.4. Spektral Eğim (dB/oktav)

Formant değerlerinin belirlenmesinde önemli bir akustik değiştirgen olan spektral eğim (*spectral tilt/slope*) (bkz. Şekil 4), konuşma spektrumunda yüksek frekanslarda en fazla –6 dB olarak hesaplanmaktadır. Spektral tepe noktaları (*spectral peaks*), gırtlak formantları (*glottal formants*) olarak tanımlanmakta, gırtlak formantının genişliği, F_0 değerine yakınlığı, yüksek frekans eğimi ve genliği gibi akustik değiştirgenlerle belirlenmektedir.



Şekil 4. "Sakar bir çocuk vazoyu düşürdü" tümcesinin spektral eğimi

F_0 değerine ya da birinci harmonik (H1) değerine çok yakın bir aralıkta konumlanan gırtlak formantı, alçak-geçirgenli filtre ile modellenmekte ve spektral eğimleri +6dB/oktav, –6dB/oktav, –6dB/oktav olmak üzere üç temel çizgisel alanda belirlenmektedir. Ses kaynağındaki düşük frekans etkisi, H1 ve H2 (ikinci harmonik) çerçevesinde gerginlik-gevşeklik (*tense-lax*) özelliklerine göre tanımlanmaktadır (Klaat & Klaat, 1990; d'Alessandro, 2012).

3. Ezgi Kuramları ve Günümüz Modelleri

3.1. Öncü Yaklaşımlar

3.1.1. İngiliz Ezgi Okulu

Hesaplamalı ve akustik sesbilim alanyazınında ezgi ve bileşenlerinin sentezlemesi üzerine öncü niteliğinde önemli yaklaşımlar bulunmaktadır. Bu yaklaşımlardan ilki olan İngiliz Ezgi Okulu (*British School of Intonation*),

YÜKSEK (*rise*) ve ALÇAK (*fall*) tonlama kavramlarını alanyazına kazandırmıştır. Palmer (1922)'nin öncülüğünde başlayan bu yaklaşım, ilerleyen dönemlerde Halliday (1967), Crystal (1969) ve O'Connor ve Arnold (1973) tarafından geliştirilmiştir. İngiliz Ezgi Okulundaki araştırmacılar, ezgi öbeği (*intonational phrase*) kavramını ve bürün sentezlemesinde (*prosody synthesis*) önemli bir yeri olan çekirdek vurgu (*nuclear accent*) kavramını da ilk kez alanyazına kazandırmıştır.

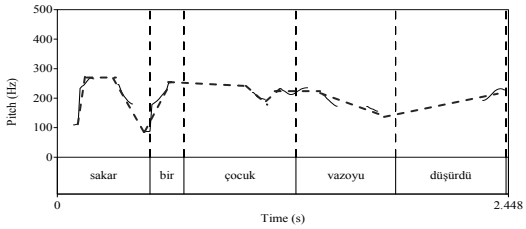
İngiliz Ezgi Okulunun en önemli araştırmacılarından biri olan Crsytal (1969), ezgi öbeğinin dildeki zorunlu bir öbek olarak kabul etmektedir. Öbeğin en önemli birimi olarak varsaydığı çekirdek, ezgi konturundaki (*contour*) başta alçak ve yüksek ton olmak üzere, alçalan-yükselen, yükselen-alçalan gibi pek çok farklı tonda yer alabilmektedir. Tonlamalara dayalı bu yaklaşım Halliday (1967)'e dayanmaktadır. Crystal'a göre, çekirdek yapının temel birimleri (1)'de sunulmaktadır. Crystal, her ezgi öbeğinin aşağıda sunulan eklenti (*tail*), baş (*head*) ve baş-öncesi (*pre-head*) olmak üzere kendi içinde bir dilbilgisel düzenlenişi olduğunu ileri sürmektedir.

(1) EKLENTİ (E): Çekirdekten sonra konumlanan birimdir.
BAŞ (B): Ezgi öbeğinin ilk vurgulu seslemi ile başlayan birimdir.
BAŞ-ÖNCESİ BİRİM (B-Ö): Çekirdeğin baş konumunu önceler.

İngiliz Ezgi Okulu, alçak ve yüksek tonlama kavramlarını sesbilim alanyazınına kazandıran ilk dilbilim okulu olmasına karşın, bu okulun özellikle akustik bileşenlere dayalı akustik bir çözümleme getirmemesi dikkat çekmiştir. Taylor (2009)'da da belirtildiği gibi, İngiliz Ezgi Okulu yalnızca öncü niteliğinde kalmış ve ezgiye dayalı çözümlemeleri yüzeysel bulunmuştur.

3.1.2. Hollanda Ezgi Okulu

Hollanda Ezgi Okulu (*Dutch School of Intonation*), ezginin perde değerinin belirlenmesinde temel sıklık (F_0) gibi önemli rolü olan akustik değiştirgenlerin incelenmesi gerektiğini vurgulamaktadır. Özellikle akustik sesbilgisine dayalı çözümlemelerinde İngiliz Ezgi Okulundan çok daha gelişmiş olan bu okul, F_0 konturlarının sadeleştirmesi (*stylization*) açısından alanyazında öncü niteliği taşımaktadır. Şekil (5)'te sunulan aşağıdaki örnekte, "*Sakar bir çocuk vazoyu düşürdü*" tümcesinin ezgi görünümü düz çizgi kullanılarak sadeleştirilmiş ve işlemlenmeyen alanlar basitleştirilmiştir.



Şekil 5. "Sakar bir çocuk vazoyu düşürdü" tümcesinin F_0 konturunda sadeleştirme görünümü

İngiliz Ezgi Okulunda çekirdeğin başı olarak tanımlanan birimler, koyu noktalı çizgilerdeki ilk yüksek perde vurgusunu göstermektedir. Hollanda Ezgi Okulu, bu alanları basitleştirmesine karşın, daha ayrıntılı olarak yorumlamakta ve biten ezgi, süren ezgi ve soru ezgisi gibi alt ezgi türleri çerçevesinde değerlendirmektedir[1]. Biten ezgi, genellikle yükselen-alçalan tonlama, süren ezgi alçalan-yükselen tonlama ve soru ezgisi ise yükselen tonlama ile gözlenmektedir. Hollanda Ezgi Okulunun öncülerinden olan t'Hart ve Cohen (1973), F_0 değerinin standardizasyonu inceleyerek, yüksek ve alçak tonlama kavramlarının yanına, yazının ilerleyen bölümlerinde ayrıntılı olarak betimlenecek olan eğim (*declination*) kavramını da eklemiştir.

3.2. Perde Değerine Dayalı Ezgi Sentezleme

3.2.1. Özerkparçalı-Ölçülü Sesbilim ve ToBI Modeli

Pierrehumbert (1980)'de sunulan doktora tezinde geliştirilen Özerkparçalı-Ölçülü Sesbilimin (*Autosegmental-Metrical Phonology*), temelleri Liberman (1975), Goldsmith (1976) ile Liberman ve Prince (1977)'e kadar dayanmaktadır. Bu kuram, günümüzde halen güncelliğini korumakta olan bir ezgi sentezlemesi içermektedir. İki parçadan oluşan Özerkparçalı-Ölçülü Sesbilim Kuramı, Goldsmith (1976)'da sunulan doktora tezinde Özerkparça Kuramı (*Autosegmental Theory*) ve Liberman (1975)'te geliştirilen Ölçülü Kuram (*Metrical Model*) ile birleştirilerek, Ladd (1996)'da Özerkparçalı-Ölçülü Sesbilim Kuramı adını almıştır.

Pierrehumbert (1980), Pierrehumbert ve Liberman (1984), Hayes (1995) ve Beckman (1996) gibi araştırmalarda özellikle İngilizce üzerinden gelişti-

[1] Biten ezgi, süren ezgi ve soru ezgisi kavramları, yükselen ve alçalan ezgiyi kullanarak, Türkçeye ilk olarak Ergenç (1989)'da kazandırılmıştır. Hollanda Ezgi Okulundaki ezgi sentezlemesi yaklaşımını benimseyen Ergenç, Türkçede bildirimin sona ermesinde alçalan tonlama ile biten ezginin, devam etmesi durumunda alçalan-yükselen tonlama ile süren ezginin ve konuşucunun dinleyiciye bir beklenti sunmasında yükselen tonlama ile soru ezgisinin kullanıldığını ileri sürmüştür.

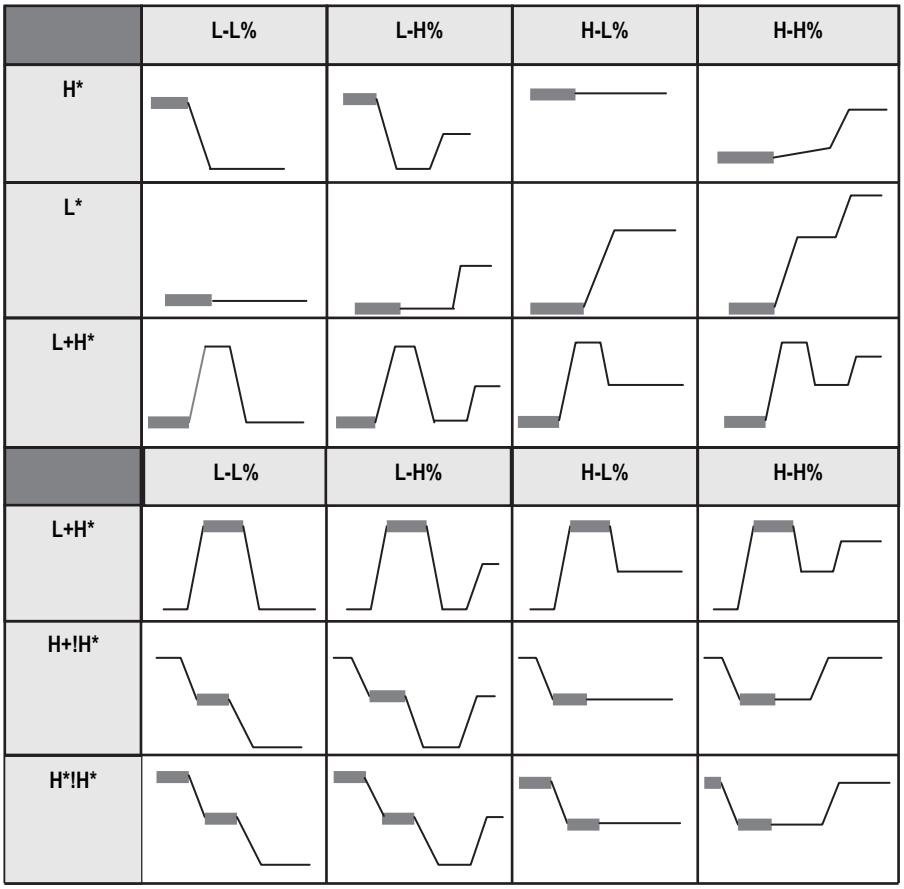
rilmiş olan bu kuram, ezgiye dayalı çözümlemeye yaparken yalnızca sesbilimden değil, akustik sesbilgisinin verilerinden de yararlanmaktadır. Bu yönüyle büyük önem kazanan bu kuram, Ladd (1996)'da Ezgisel Sesbilim (*Intonational Phonology*) olarak da tanımlanmaktadır. Bilindiği gibi, İngiliz Ezgi Okulunda ezgi öbeğine dayalı bir çözümleme yalnızca çekirdek ve başın etrafında sesbilimsel verilere dayalı olarak çözümlenmektedir. Ancak Özerkparçalı-Ölçülü Sesbilim Kuramı, çekirdek ve baş kavramları arasındaki farklılığı ortadan kaldırarak, öbek sınırlarına ve olası perde vurgularına dayalı yeni bir çözümlemeyi alanyazına kazandırmıştır.

Özerkparçalı-Ölçülü Sesbilim Kuramındaki akustik çözümlemeye göre, olası bir perde vurgusunda bir yüksek (*high*) ve alçak (*low*) tonlama bulunmakta ve bu tonlamalar çeşitli perde vurgusu düzenleri çerçevesinde dillerarasında farklı biçimlerde yorumlanmaktadır. İngilizce temelinde hazırlanan bu ton işaretleme sistemi Pierrehumbert (1980)'de ToBI (*Tones and Break Indices*) olarak tanımlanmaktadır. ToBI sisteminde yıldız işaretli ton **(*)** (*starred tone*) vurgulu seslemi işaretlerken, sınır işaretli ton **(%)** (*boundary tones*) öbek sınırlarını, öbek işaretli ton **(-)** (*phrase tones*) çekirdek vurgudan öbek sınırına geçişi işaretlemektedir. ToBI, günümüzde Türkçe dahil olmak üzere (bkz. Kan, 2009; Kamali, 2011; Güneş, 2015; İpek, 2015) pek çok farklı sesbilimsel görünüm içeren dilde yorumlanabilir bir tonlama sistemidir. Tablo (1) ve Tablo (2)'de ToBI'ye dayalı olarak Pierrehumbert (1980), Pierrehumbert ve Liberman (1984), Beckman (1996) gibi araştırmalar çerçevesinde özetlenebilir yükselen ve alçalan tonlama düzenleri ile ezgi öbeğindeki olası perde vurgularının genel hiyerarşisi sunulmaktadır.

Tablo 1. ToBI'ye dayalı tonlama düzenleri

Tek Tonlar	**Tonlama Düzenleri**
H*	Vurgulu seslemdeki en yüksek perde vurgusudur.
L*	Vurgulu seslemdeki en düşük perde vurgusudur.
Çoklu Tonlar	**Tonlama Düzenleri**
H+L*	Vurgulu seslemdeki perdenin en yüksek noktasıdır.
L+H*	Vurgulu seslemdeki düşükten yüksek tona geçiştir.
Sınır Tonları	**Tonlama Düzenleri**
L-	Perdenin en düşük temel sıklık (F_0) noktasıdır.
L-L%	Alçak tondan sonraki en düşük sınır tonlamasıdır.
H-	Perde vurgusunu yükselten tondur.
H-H%	Yüksek tondan sonraki en yüksek sınır tonlamasıdır.
H-L%	Yüksek tondan sonraki alçak sınır tonlamasıdır.

Tablo 2. ToBI'ye dayalı perde vurguları



Türkçede Özerkparçalı-Ölçülü Sesbilim Kuramı temelinde oluşturulmuş öncü çalışmalardan biri olan Kan (2009)'da Türkçedeki perde vurguları için (H*), (!H*), (L+H*) ve (L+!H*) ve öbeksel sınırda (L+H-) ya da (L+!H-) tonlama düzeni önerilmiştir. Türkçede sözcük vurgusunu Özerkparçalı-Ölçülü Sesbilim Kuramı temelinde inceleyen Kamali (2011)'de Türkçe sözlüksel olarak perde vurgusu dili olarak kabul edilmiş, vurgunun sonda olmadığı sözcüklerde (H*+L) tonlama düzeninin bulunduğu ileri sürülmüştür. Kamali'ye göre, Türkçede sözcük vurgusu serbest bir düzenleniş içermektedir. İpek (2015)'te ise, Türkçede perde vurgusu ToBI çeviriyazı sistemi üzerinden incelenerek, vurgu düzenlemesinde öbeksel belirginlik (*phrasal prominence*), kenar işaretli sınır tonlaması ile baş işaretli perde vurgusunun önemli rol oynadığı belirtilmiştir.

AM Modeli olarak da tanımlanabilir olan Özerkparçalı-Ölçülü Sesbilim Kuramında eğime dayalı yorumlamalar, F_0 konturundaki düzeye dayalı tonlama (*downdrift*) ile ilişkilendirilmektedir. Taylor (2009)'a göre bu örüntü, perde vurgusuyla aynı düzeyde yukarıdan aşağıya iniş (*downstep*) etkisinden kaynaklanmaktadır. Bu etki, çoğunlukla konuşucu tarafından da kontrol edilebilmektedir. Pierrehumbert (1980), tonlamada yukarıdan aşağıya iniş etkisinin özellikle Afrika ton dillerinde açıkça gözlemlenebilir olduğunu ileri sürmektedir.

3.2.2. INTSINT Ezgi Modeli

Fransız sesbilimci Daniel Hirst tarafından 1987'de geliştirilen Intsint Ezgi Modeli (***In****ternational* ***T****ranscription* ***S****ystem for* ***Int****onation*), Uluslararası Sesbilim Abecesinin (*International Phonetic Alphabet*) bürünsel karşılığı olarak ortaya koyulmuş ve çokdilli bir çeviriyazı oluşturulmaya çalışılmıştır (bkz. Hirst & Di Cristo, 1998; Hirst, Di Cristo & Espesser, 2000; Hirst, 2001). Taylor (2009)'da kuramdan bağımsız bir model olarak tanımlanan Intsint Ezgi Modeli, Uluslararası Sesbilim Abecesine dayalı bürün etiketleme sisteminde yandaki kodlamaları kullanmaktadır. Ses yükseklik düzeyi (*pitch range*), **T**, **M** ve **B**, **H**, **S** ve **L** ile belirlenirken, **D** ve **U** vurguya dayalı hedef noktalarını işaretlemektedir.

(2)

T (en yüksek) (*top*)
H (daha yüksek) (*higher*)
U (kademeli yüksek) (*upstepped*)
S (aynı düzey) (*same*)
M (orta düzey) (*mid*)
D (yukarıdan aşağıya iniş) (*downstepped*)
L (daha alçak) (*lower*)
B (dip) (*bottom*)

Intsint Ezgi Modelindeki akustik çözümleme, konuşmacı ve dinleyici arasındaki sesbilgisel değişkenlere dayalı olarak da yorumlanmaktadır. Hirst (2004), modelde anahtar (*key*) ve düzen (*range*) olarak tanımladığı iki temel değiştirgenden söz etmektedir. Anahtar, F_0 (temel sıklık) değerinin belirgin noktasını gösterirken, düzen sözcedeki en yüksek ve en düşük perde değerleri arasındaki uzaklığı belirlemektedir. Hirst'te sunulan algoritma aşağıdaki gibidir:

(3)

T: P(i) : = anahtar + düzen/2
M: P(i) : = anahtar
B: P(i) : = anahtar - düzen/2

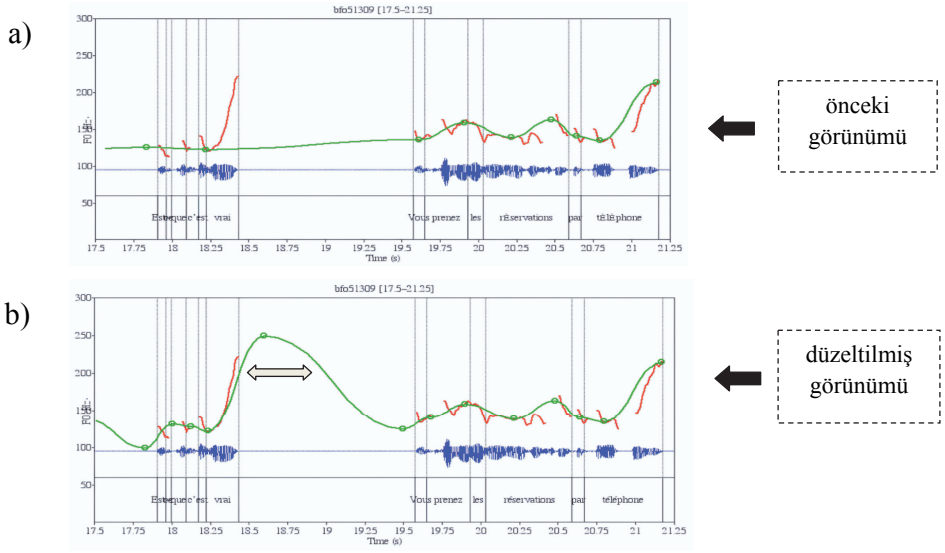
H: P(i) := (P(i-1) + T) / 2
U: P(i) := (3*P(i-1) + T) / 4
S: P(i) := P(i-1)
D: P(i) := (3*P(i-1) + B) / 4
L: P(i) := (P(i-1) + B) / 2

(3)'teki algoritmada genellikle logaritmik ölçeğe dayalı olarak hesaplanmakta olan perde değerlerinde, yukarıdaki formüllerde görülen **P(i)** o anki perde vurgusuna ait hedefi, **P(i-1)** ise önceki perde vurgusuna ilişkin hedefi göstermektedir.

Hirst (2004), Intsint Ezgi Modelinin, Uluslararası Sesbilim Abecesinde olduğu gibi, çok dilli bir sistem üzerinden bürünsel kodlamalarla hesaplanabilir olduğunu ve her dil için ayrı bir etiketleme sistemi oluşturulabileceğini ileri sürmektedir. Hesaplamalı sesbilgisi alanyazınına önemli bir katkı sunan bu model, yalnızca akustik sesbilgisel düzlemde kaldığı için ve sesbilimsel yorumlamalardan uzak bir hesaplama düzeni oluşturduğu için sesbilimciler tarafından çok fazla benimsenmemiştir. Ancak yine de, akustik ve hesaplamalı sesbilgisi alanlarının arakesitinde geliştirilmiş önemli bir etiketleme sistemi olarak kabul edilmesi gerekmektedir.

3.2.3. MOMEL Ezgi Modeli

Daniel Hirst tarafından 1980'de geliştirilmiş olan ve bir önceki bölümde betimlenen Intsint Ezgi Modeli ile ilişkilendirilen Momel (*Modélisation de MELodie*) Ezgi Modeli, Hirst ve Espessser (1993)'te özdevinimleştirilmiştir. Momel algoritmasının temel özelliği hedef noktalarını (*target points*) sıklık ve süre eğrisinde F_0 değerleri çerçevesinde hesaplamaktadır. Hirst, Intsint Ezgi Modelinde geliştirilmiş olduğu evrensel yaklaşımı, Momel Ezgi Modeli algoritmasında da sürdürmüştür. Model İngilizce, İspanyolca, Arapça, İtalyanca ve Fransızca gibi pek çok dilde F_0 örüntüsünü belirleme amacıyla uygulanmıştır (Ayrıntılı bilgi için bkz. Campione, Flachaire, Hirst, & Véronis, 1997; Campione & Véronis, 1998; Hirst, Ide, & Véronis, 1994, Hirst, Di Cristo, & Espesser, 2000). Aşağıda Hirst (2007: 1235)'den alıntılanarak uyarlanan Şekil (6)'da Momel Ezgi Modelinin farklı görünümleri sunulmaktadır.



Şekil 6. MOMEL Ezgi Modeli

Fransızca "Est-ce que c'est vrai? vous prenez les réservations par téléphone?" (*Bu doğru mu? Rezervasyonları telefonla mı alıyorsunuz?*) tümcesinin F_0 perde örüntüsündeki parçalı (kırmızı) ezgi eğrileri, Praat (Boersma & Weenink, 2020) ses analizi yazılımından alınan ham veriyi (İng. *raw data*) alanlarını göstermektedir. Düzeltilmiş algoritmada birleştirilmiş (yeşil) ezgi eğrileri ise modellenmiş, başka bir deyişle işlenmiş veriyi sunmaktadır. Bu açıdan, Şekil (6)'da önceki F_0 örüntüsünde yükselen tonlama değerleri hedef noktalarının dışında hesaplanırken, düzeltilmiş görünümde F_0 örüntüsü üzerindeki tüm tonlama değerleri tek tek hedef noktaları biçiminde kabul edilmiştir. Bu durum, ezgi eğrilerine ilişkin tüm tonlama değerlerinin dış değerlerine ilişkin içbükey alanlarının ve yakınlık içeren alanların tahmini olarak hesaplanmasıdır.

Bu yönleriyle, Hirst (2007)'de de belirtildiği gibi, Momel Ezgi Modeli algoritmasının aslında Intsint Ezgi Modeli içinde bütünleştirilmiş bir model olarak kabul edilmesi ve bu iki modelin birlikte ham veri üzerinden işlenmesi büyük önem taşımaktadır. Bu amaçla, Hirst günümüzde sesbilim alanında araştırmalarını sürdürmekte olan araştırmacıların yoğun olarak kullandığı Praat ses analizi yazılımında eklenti komut dizisi (*plugin*) oluşturmuştur. Bu eklenti komut dizisi aracılığıyla, ezgi eğrilerinde F_0 örüntüsünün ve Momel hedef noktalarının belirlemesi ile Intsint algoritmasına ilişkin kodlamaların yapılabilmesi mümkün hale gelmektedir.

3.2.4. Fujisaki Üstdüşümsel Modeli

Fujisaki ve Kawai (1982)'de sunulan Fujisaki Üstdüşümsel Modeli (*Fujisaki's Superpositional Model*), ezgi sentezlemesinde öbek (*phrase*) ve vurgu (*accent*) olmak üzere iki temel kavram üzerinden durmaktadır. Hesaplamalı sesbilgisi alanyazında önemli bir yeri olan Fujisaki Üstdüşümsel Modeli, F_0 konturlarının sözlü dil üretim mekanizmasındaki yansımasını incelemekte ve Taylor (2009: 242)'dan uyarlanan aşağıdaki Şekil (7)'deki algoritma üzerinden hesaplama yapmaktadır.

(4)

$$\ln F_0(t) = \ln F_{min} + \sum_{i=1}^{I} A_{p_i} G_{p_i}(t - T_{0_i}) + \sum_{j=1}^{J} A_{a_j}(G_{a_j}(t - T_{1_j}) - G_{a_j}(t - T_{2_j}))$$

$$G_{p_i}(t) = \begin{cases} \alpha_i^2 t e^{-\alpha_i t} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases} \qquad \textit{olduğunda}$$

ve

$$G_{a_j}(t) = \begin{cases} \min[1 - (1 + \beta_j t)e^{-\beta_j t}, \theta] & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

Şekil 7. Fujisaki Üstdüşümsel Modelinin temel algoritması

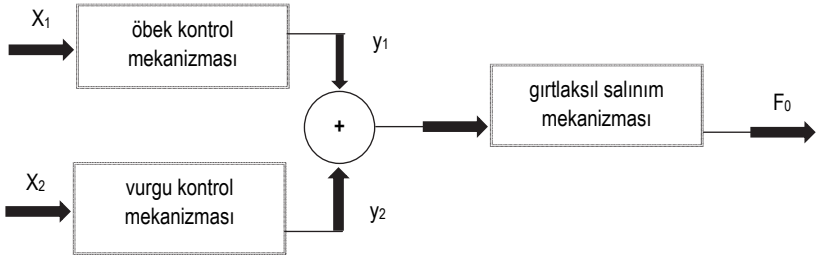
Yukarıda sunulan bu algoritmada sönümlü FIR süzgeci (*finite impulse filters*) çerçevesinde sıralı bir mekanizmayı göstermektedir. Buna göre, Taylor (2009)'da da belirtildiği gibi, süzgeçlerden biri öbek bileşeninin, diğeri ise vurgu bileşeninin hesaplanmasında etken rol oynamaktadır. Algoritmadaki her fonksiyonun işlevsel karşılığı aşağıdaki gibidir:

(5)

F_{min}: temel hat/taban
I: öbek bileşenlerinin sayısı
J: vurgu bileşenlerinin sayısı
A_{pi}: $_i$'nin öbek büyüklüğü
A_{aj}: $_j$'nin vurgu büyüklüğü
T_{0i}: $_i$'nin öbek komut zamanlaması
T_{1j}: $_j$'nin vurgu komut başlangıcı
T_{2j}: $_j$'nin vurgu komut bitişi
α_i: $_i$ öbek komutunun öbek kontrol mekanizmasının doğal açısal sıklığı
β_j: $_j$ vurgu komutunun vurgu kontrol mekanizmasının doğal açısal sıklığı
θ: vurgu bileşeninin yükseklik sınırını gösteren değiştirgeni

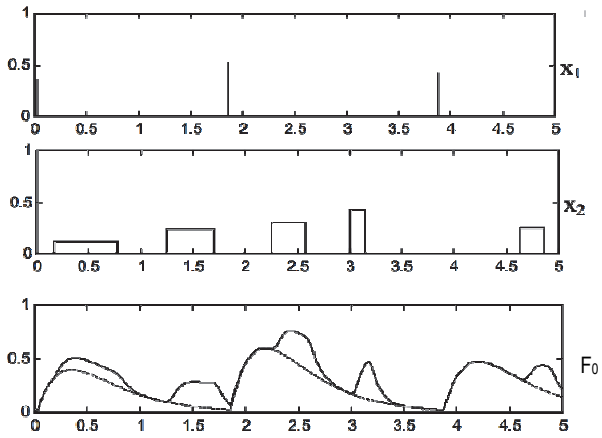
Fujisaki Üstdüşümsel Modelinde öbek ve vurgunun birlikte işlemlenmesine ve hesaplanmasına yönelik algoritma, Rossi, Palmieri ve

Cutugno (2002)'de aşağıdaki Şekil (8)'deki gibi basitleştirilerek, F_0'a erişim süreciyle ilişkilendirilmiştir. Şekilde sunulan X_1 alanı öbek-komut girdisini (*phrase-command input*), X_2 alanı vurgu-komut girdisini (*accent-command input*), y_1 alanı F_0 değerine erişim için öbek-komut girdisini (*phrase-command contribute to F_0*) ve y_2 alanı ise F_0 değerine erişim için vurgu-komut girdisini (*accent-command contribute to F_0*) tanımlamaktadır.



Şekil 8. Fujisaki Üstdüşümsel Modelinde öbek-vurgu kontrol mekanizması

Rossi, Palmieri ve Cutugno (2002: 1)'den uyarlanan aşağıdaki şekillerdeki F_0 konturlarında (bkz. Şekil 9), X_1 alanındaki öbek-komut girdisinin öbek-kontrol mekanizmasıyla ve X_2 alanındaki vurgu-komut girdisinin vurgu komut mekanizmasıyla olan bağıntısı gösterilmektedir. Bask dili (bkz. Navas, Hernaez, & Sanchez, 2000) ve İsveçce (bkz. Fujisaki, Ljungqvist, & Murata, 1993) gibi pek çok dilde sınanmış olan Fujisaki Üstdüşümsel Modeli, F_0 örüntüsünde yukarıda da söz edildiği gibi X_1 ve X_2 alanlarının hesaplanması üzerine kurulu bir sistemi içermektedir. Şekil (9)'da F_0 değerinin hesaplanması üzerine kapsamlı Fujisaki örüntüsü gösterilmektedir.

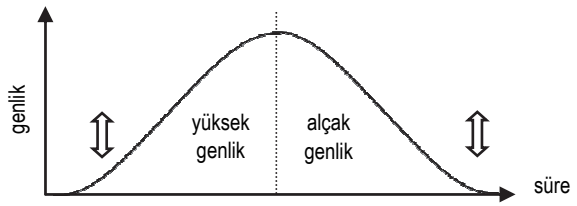


Şekil 9. Fujisaki Üstdüşümsel Modelinde F_0 örüntüsünün açılımı

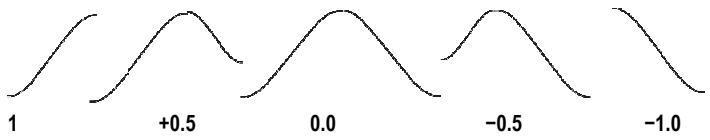
Fujisaki Üstdüşümsel Modeli pek çok algoritma içerdiği için oldukça karmaşık gibi görünüyor olsa da, aslında temel mekanizması oldukça sadedir. Modelde yalnızca öbek ve vurgu arasındaki temel ilişki sınanmaktadır. Her bir kontrol mekanizmasının süzgeç aşamasına gelmeden önce tek bir temel bir etkisi bulunmaktadır. Taylor (2009)'da da tartışıldığı gibi, vurgunun hesaplanmasında, süzgece verilen yanıt hızı önem taşımaktadır. Buna göre, genellikle öbek bileşeni vurgu bileşeninden hem ilk yanıtta hem de düşüş aşamalarında daha hızlı bir örüntü sergilemektedir. Sade bir model olmasına karşın Fujisaki Üstdüşümsel Modeli, sesbilimsel değişkenleri ve insana özgü değişmezleri ve bireysel faktörleri çok fazla temel almadığı için fazla somut bir model olarak kabul edilmektedir.

3.2.5. Eğim (TILT) Modeli

Taylor (1992)'de geliştirilen ve ezgi sentezlemesi alanyazında önemli bir yeri olan Eğim Modeli (*Tilt Model*) hesaplamalı bir akustik modeldir. Özerkparçalı-Ölçülü Sesbilim Kuramı üzerinden geliştirilmiş olan ve sesbilimsel açıdan bu kuramı temel alan Eğim Modeli, ezgiyi olay dizileri (*event series*) olarak hesaplamaktadır. Bu yönüyle Eğim Modeli, bir önceki bölümde betimlenen Fujisaki Üstdüşümsel Modelinden farklılaşmaktadır. Eğim Modeli, insana özgü değişmezleri de sesbilimsel görünümler çerçevesinde ele almaktadır. Bu model, aynı zamanda dillere özgü F_0 değerlerinin ve perde örüntülerinin belirlenmesinde karmaşık olmayan ve sıralı bir sentezleme algoritması sunmaktadır. Özerkparçalı-Ölçülü Sesbilim Kuramının temel perde ölçümleri olan yüksek ve alçak tonlama değerleri, Taylor (1992)'den uyarlanan aşağıdaki Şekil (10) ve Şekil (11)'de sunulduğu gibi, Eğim Modelinde de en temel bileşenlerdir.



Şekil 10. F_0 perde eğrisinde genlik ve süre grafiği



Şekil 11. Eğim değerleri [+1 ile −1] arasındaki perde vurguları

Eğim Modeli, yukarıdaki şekillerde de görüldüğü gibi, perde değerlerini Özerkparçalı-Ölçülü Sesbilim Kuramından farklı olarak yüksek ve alçak tonlama belirlemesinin yanında, bu değerlere perde görünümlerindeki eğim düzeylerine göre sayısal bir değer atamaktadır. Olay dizileri olarak da bilinen bu atamalar, perdenin yüksekliğine, öbek sınırlarına ve perdenin alçalmasına dayalı olarak [+1 ile −1] arasında değer almaktadır. En uç noktalar [+1] olarak saf yükseklik (*pure rise*) ile [−1] olarak saf alçaklık (*pure fall*) biçiminde kodlanmaktadır. Perdedeki iki eğrinin kesiştiği alan ise sıfır geçiş noktası olarak [0.0] kabul edilmektedir.

Fujisaki Üstdüşümsel Modelinden farklı olarak Eğim Modelinde vurgu ve sınır (*boundary*) olmak üzere iki temel olay dizisi bulunmaktadır. Bu olay dizileri, Özerkparçalı-Ölçülü Sesbilim Kuramında olduğu gibi, perde vurgularındaki değişimlere dayalı olarak alçalan ve yükselen tonlamalar çerçevesinde yorumlanmaktadır. Şekil (11)'deki sıfır geçiş noktasında gözlemlenen [0.0] eğriler arası birleşim yerleri ise bağlantı noktası (*connections*) olarak tanımlanmaktadır. Taylor (2009)'a göre, perde örüntüsünün ölçümünde olay dizisinin şeklini inceleyen dört değiştirgen ve bağlantı sırasını inceleyen iki temel değiştirgen bulunmaktadır. Bu modelleme Taylor (1995)'te önerilen Yükselen/Alçalan/Bağlantı Modeli (*Rise/Fall/Connection Model*) olarak da adlandırılmaktadır. Şekil (10)'daki F_0 perde eğrisi grafiğinde de görüldüğü gibi, Eğim Modeli her bir olay dizisi için dört temel değiştirgen ve üç eğim değiştirgeni içermektedir. Genlik (*amplitude*), süre (*duration*) ve eğim (*tilt*) çerçevesinde oluşturan Eğim Modelindeki temel perde örüntüsü algoritması (6)'daki gibidir:

(6)

$$\text{Genlik} = G = |G_{yükselen}| + |G_{alçalan}| \qquad Eğim_{genlik} = \frac{|G_{yükselen}| - |G_{alçalan}|}{|G_{yükselen}| + |G_{alçalan}|}$$

$$\text{Süre} = S = |S_{yükselen}| + |S_{alçalan}| \qquad Eğim_{süre} = \frac{|D_{yükselen}| - |D_{alçalan}|}{|D_{yükselen}| + |D_{alçalan}|}$$

$$\textbf{EğimDeğişkeni} \quad eğim = \frac{eğim_{genlik} + eğim_{süre}}{2}$$

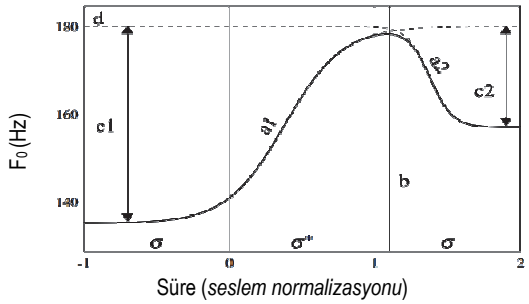
Yukarıda sunulan algoritmalardan da anlaşılacağı gibi, öncelikli olarak toplam genlik değerinin belirlenmesi için yükselen ve alçalan genlik değerleri toplanmaktadır. Aynı formülleştirme, süre değerinin belirlenmesinde de

kullanılmaktadır. Bunun ardından, hem eğimin genlik değeri hem de eğimin süre değerinin belirlenmesi için yükselen-alçalan tonlama değerlerinin çıkarım ve toplam değerleri birbirine bölünmektedir. Bu işlemle ortalama bir eğim değeri (genlik ve süre) elde edilmektedir. Eğim değişkeninin belirlendiği en son aşamada ise, eğimin genlik değeri ile eğimin süre değerleri toplanıp, ikiye bölünmektedir. Bu işlemden sonra, perde örüntüsünün eğim hesaplaması tamamlanmaktadır.

Eğim Modeli, Özerkparçalı-Ölçülü Sesbilim Kuramının geliştirilmiş ve hesaplamalı akustik bir modeli olarak kabul edilebilmektedir. Bu yönleriyle, kişiye özgü değişmezleri temel alması ve bireysel farklılıklar çerçevesinde bir hesaplama sistemi oluşturması açısından önemli bir akustik modeldir. Ancak, yine de sesbilimsel değiştirgenleri hesaplarken soyutluk öte, somutluğu temel aldığı için eksik kaldığı noktalar bulunmaktadır. Sesin fiziğe ait somut bir kavram olduğunu düşünsek bile, Eğim Modeli pek çok yönüyle geliştirilmiş bir akustik model olmasına karşın, F_0 perde örüntüsünde ezginin belirlenmesinde ses üretimindeki değişimlere yönelik sesbilimsel etkenleri daha fazla sorgulaması beklenmektedir.

3.2.6. PAintE Ezgi Modeli

Möhler ve Conkie (1998)'de ortaya konulup, Möhler (2001)'de geliştirilmiş olan PAintE Ezgi Modeli (***Pa****rametric Representations of* ***Int****onational* ***E****vents*), Şekil (12)'de sunulan s-biçimindeki (*sigmoid*) ezgi eğrisine ait yaklaşık yükselen ve alçalan tonlama değerlerini parametrelerle belirlemektedir. PAintE Ezgi Modelinde her bir parametreyi seslemde yüksek tonlama içeren (σ∗) vurgulu seslem görünümleri ile önceki ve sonraki seslemlerle çerçevesinde incelemektedir. Her bir seslemin uzunluğu (1)'e normalize edilmekte, böylelikle seslemler −1 ile 2 aralığında varsayılmaktadır. Aşağıda sunulan PAintE Ezgi Modelinde altı temel değiştirgen bulunmaktadır. Buna göre, ***c1*** ve ***a1*** yüksek tonlamadaki s-biçimin yüksekliği ve eğrisini, ***c2*** ve ***a2*** alçalan tonlamadaki s-biçimin yüksekliği ve eğrisini, ***d*** ve ***b*** ise belirgin yükseklik ve ezgi olayındaki ilgili seslem hizalamasını belirlemektedir. Şekil (12)'de Schweitzer, Walsh ve Schweitzer (2017)'den uyarlanan PAintE Ezgi Modelinin genel F_0 perde örüntüsü gösterilmektedir.



Şekil 12. PAintE Ezgi Modelinin genel F_0 perde örüntüsü

Yukarıda sunulan s-biçimli ezgi eğrisinde eğer vurgulu seslem, iki s-biçim arasına yerleştirilemezse, yalnızca tek bir s-biçim alanı kullanılmaktadır. Bu durumda, Gessinger, Schweitzer, Andreeva, Raveh, Möbius ve Steiner (2018)'de belirtildiği gibi, ***c*** ve ***a*** değiştirgenlerinin bir kısmı ezgi alanı dışında bırakılmaktadır. Ancak, eğer ki tek bir s-biçim iyi bir seçim olarak kabul edilmezse, bu durumda da modelde yalnızca F_0 perde örüntüsünün ortalama değeri ***d*** alanı olarak varsayılmakta, diğer alanları analiz dışında bırakılmaktadır. Möhler ve Conkie (1998)'de PAintE Ezgi Modelinde iki s-biçimi arasındaki algoritma (7)'de sunulmaktadır.

(7)

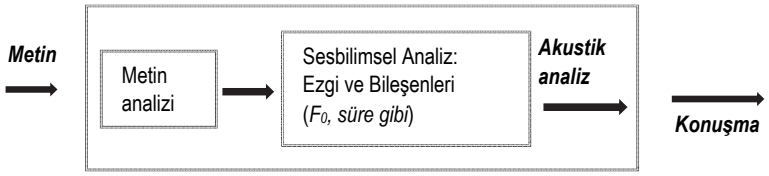
$$f(x) = \frac{c_1}{1 + \exp(-a_1 b - x) + \gamma)} - \frac{c_2}{1 + \exp(-a_2 x - b) + \gamma)}$$

Algoritmada a_1 ve a_2 alanları s-biçimlerin yüksekliğini, c_1 ve c_2 alanları s-biçimlerin genliğini, ***b*** alanı hizalama kısmını ve ***d*** ise tepe noktasının işlevini işaretlemektedir. Öztürk (2006)'da bu yönleriyle, aslında PAintE Ezgi Modeli, bir önceki bölümde sözü edilen Eğim Modeli ile pek çok noktada benzerlik taşımaktadır. Her iki modelde de, ezgi eğrileri vurgulu seslemleri referans noktaları alan ezgi olayları çerçevesinde hesaplanmaktadır. PAintE, günümüzde ezgi eğrilerinin F_0 değerlerinin hesaplanması açısından en pratik modellerden biri olarak kabul edilmektedir.

4. Metinden Konuşma Sentezleme (MKS): Türkçede Ezgi ve Bileşenleri Modellemesi

Türkçede ezgi ve bileşenlerinin hesaplamalı sesbilgisi alanındaki gözlem teknikleri ışığında modellemesinin yapıldığı araştırmalar, genellikle mühendislik alanlarında artış göstermektedir. Günümüze kadar gerçekleştiri-

len araştırmalar, konuşmacı tanıma/tanımlama, Metinden Konuşma Sentezleme (MKS) (*text-to-speech synthesis/TTS*), bürün modellemesi (*prosodic modelling*) gibi konuları kapsamaktadır. Kapsamlı bir araştırma alanının konusu olan Metinden Konuşma Sentezleme, bu bölümün sınırlılıkları çerçevesinde Şekil (13)'te gibi bir modelleme ile özetlenebilir.



Şekil 13. Metinden Konuşma Sentezlemenin Genel Görünümü

Allen, Hunnicut ve Klatt (1987)'de alanyazına kazandırılan ve günümüzde oldukça geniş bir araştırma alanı taşıyan Metinden Konuşma Sentezleme, özellikle hesaplamalı sesbilgisi alanında önemli bir yer edinmiştir. Bilindiği gibi, konuşma sentezleme insan sesinin yapay bir üretimini sunmaktadır. Yazılı bir metnin bilgisayar ortamında sözlü üretime dönüştürülmesi ve bu süreçte mümkün olduğunca doğal bir üretim oluşturulmaya çalışılmaktadır. Metinden Konuşma Sentezleme sürecinde pek çok alt sentezleme sistemi bulunmaktadır:

a. Eklemeli Konuşma Sentezleme (Hunt & Black, 1996)
(*Concatenative Approach to Text-to-Speech Synthesis*)

b. Birim Seçimi Tabanlı Konuşma Sentezleme (Hunt & Black, 1996)
(*Unit Selection Speech Synthesis*)

c. İstatistiksel Konuşma Sentezleme (Zen, Tokuda & Black, 2009)
(*Statistical Parametric Speech Synthesis*)

d. Hidden Markov Tabanlı Konuşma Sentezleme (Yoshimura, Tokuda, Masuko, Kobayashi & Kitamura, 1999)
(*Hidden Markov Model based Text-to-Speech Synthesis*)

Yukarıdaki alt sentezleme sistemlerinin her biri kullandığı veri tabanlarının niteliklerine dayalı olarak uygulamada kolaylık ya da zorluk içermektedir. Bu bağlamda, kısaca söz etmek gerekirse, Birim Seçimi Tabanlı Konuşma Sentezleme sistemi mümkün olduğunca doğal ve büyük veri tabanlarını kabul ederken, Hidden Markov Tabanlı Konuşma Sentezleme sistemi çok daha az doğallık içermesine karşın, daha küçük veri tabanlarıyla çalışılmasını mümkün kılmaktadır.

Türkçede ezginin modellenmesi üzerine yapılmış araştırmalar arasında öncü niteliğini taşıyan Oskay (2002)'de Türkçenin bürün dizgesinin özdevinimli modelleme (*Automatic Modelling of Turkish Prosody*) denemesi gerçekleştirilmiştir. Bu tez çalışmasıyla da ilişkili Oskay, Salor, Özkan, Demirekler ve Çiloğlu (2001)'de, olumlu ya da olumsuz tümcelerde ezginin işleyişi, F_0 değerlerindeki doğrusal ya da ikinci dereceden fonksiyonların kullanımına yönelik bir modelleme geliştirilmiştir.

Ezginin süre bileşenin üzerinde inceleme yapan Şayli (2002)'de ve Vural (2003)'te MKS sistemi kullanılarak, Türkçedeki sesbirimlerin süre özellikleri ortaya konulmuştur. Benzer şekilde sesbirimlerin süre özelliklerini hesaplayarak ezgi modellemesi geliştiren Öztürk (2006)'daki doktora tez çalışmasında ise, sesbirimlerin hem süre hem de F_0 değerleri içn modelleme oluşturulmuştur. Bu açıdan oldukça dikkat çekici bir araştırma olan Öztürk'te sesbirimlerin görünümleri, farklılıkları, sözcükteki seslem sayısı, seslemin konumlanışı, niteliği (vurgulu olup olmaması) gibi pek çok sesbilimsel etken dikkate alınmıştır. Abdullahbeşe (2011)'de ise Türkçede MKS sisteminde benzer yaklaşımlar temel alınarak, ezginin F_0 bileşeni üzerinden yeni bir ezgi modellemesi geliştirilmeye çalışılmıştır.

Türk (2003), Türk (2007) ve Salor (2005)'teki tez çalışmalarında, MKS sistemi üzerinden ses dönüştürme (*voice conversion*) ve Türkçede konuşma analizinin ilerlemesi için araç geliştirme üzerine yeni teknikler geliştirilmeye çalışılmıştır. Bir diğer önemli yenilik, Külekçi ve Oflazer (2006)'daki araştırmada gözlenmiştir. Araştırmacılar, Türkçede öbek vurgusu içeren söz öbeklerindeki ezgi görünümlerine derece ekleyerek, (0: yok), (1: az) ve (2: fazla) gibi sınıflandırma geliştirerek, bu sınıflandırmanın MKS ile birleştirilmesinin önemini vurgulamışlardır. Uslu ve İlk (2012)'de ise kural-tabanlı duyumsal bir ezgi modeli geliştirilerek, tümce düzeyinde öbek vurgusu taşıyan sözcükler, çekimli eylemler ve kimi noktalama imleri üzerinde ezgi sentezlemesi yapılmıştır. Araştırmada, 16 kHz örnekleme oranında kullanılan ikili ses (*diphone*) ile ilişkilendirilmiş ezgi modeli uygulanmıştır. Ham verilerle karşılaştırıldığında uygulanan ezgi modelinin daha iyi sonuçlar verdiği bulgusuna ulaşılmıştır.

Konuşmacı tanıma/tanımlama üzerine de Türkçede dikkate değer nitelikli ezgi modelleme sistemleri oluşturulduğu görülmektedir. Bu konuda yapılmış en kapsamlı tez çalışmalarından biri olan Dalva (2012)'de, özdevinimli konuşmacı tanıma sistemi geliştirilerek, Türkçede bürünsel özellikleri gösteren işaretler kullanılarak, tümce bölütlemesi yapılabilen güncel bir sis-

tem ortaya konulmuştur. Bu araştırma temelinde geliştirilen Dalva, Revidi, Güz ve Gürkan (2014)'te, bürünsel bilginin işaretlenmesi üzerine günümüz haber metinleri incelenmiş ve sözcük dizilerinin oluşturulabilmesinde Hidden Markov Tabanlı Konuşma Sentezleme sisteminden yararlanılmıştır. Araştırmada, bürünsel işaretlemede etkin rol olarak temel ezgi bileşenleri araştırılmış ve başarım oranları hesaplanmıştır ve süre > temel sıklık > enerji yoğunluğu üzerinden başarım oranları elde edilmiştir.

5. Bitirirken

*5. Hesaplamalı Sesbilgisi*nin bu bölümünde, alanyazında önemli bir yeri olan ezgi sentezleme ve Metinden Konuşma Sentezleme alanları üzerine oluşturulmuş kuramsal modeller ve Türkçe üzerine yapılmış araştırmalar ana çizgileriyle betimlenmeye çalışılmıştır. Bölümün ilk kısmında, hesaplamalı sesbilgisi alanı ve akustik sesbilgisi alanları arasındaki etkileşim tanımlanmış, ikinci kısmında ise ezgi ve ezgiyi oluşturan bileşenlere özgü temel sıklık (F_0), birinci ve ikinci formant sıklıkları (F_1 ve F_2), enerji yoğunluğu ve spektral eğim gibi akustik tabanlı değişkenlerinin temel özellikleri açıklanmıştır.

Hesaplamalı sesbilgisi ve akustik sesbilgisi alanyazınlarında, bürün dizgesinin işleyişinde önemli bir inceleme alanı oluşturan ezgi, günümüze kadar sesbilimsel ve sesbilgisi tabanlı pek çok araştırmada ve kuramsal yaklaşımda ele alınmıştır. Bu bölümün üçüncü kısmında İngiliz Ezgi Okulu (Halliday, 1967; Crystal, 1969; O'Connor & Arnold, 1973) ve Hollanda Ezgi Okulu (t'Hart & Cohen, 1973) gibi ezgideki yükselen ve alçalan ton değişimlerini ele alan öncü kuramsal yaklaşımlardan söz edilmiştir. Hesaplamalı sesbilgisi alanının önünü açan bu kuramsal okullar, üçüncü bölümün ikinci kısmında sözü edilmiş olan ezgi sentezleme modellerinin de tabanını oluşturmaktadır. Bu bölümde hesaplamalı sesbilgisi alanyazınına dayandırılarak betimlenmeye çalışılan kuramsal yaklaşımlar ve modeller sırasıyla şu şekilde özetlenebilir:

a. *Özerkparçalı-Ölçülü Sesbilim ve ToBI Modeli* (Liberman, 1975; Goldsmith, 1976; Liberman & Prince, 1977; Pierrehumbert, 1980; Ladd, 1996), ezgiyi yükselen ve alçalan tonlamalar çerçevesinde inceleyerek, İngilizceye özgü tonlama düzenleri oluşturmuştur. Dönemine göre ezginin ilk kez bu kadar ayrıntılı incelendiği bu kuramsal yaklaşım ve model, ezgi sistemini tek, çoklu ve sınır tonlamalarından oluşan bir tonlama düzeni olarak varsaymaktadır.

b. *Intsint Ezgi Modeli* (Hirst & Di Cristo, 1998; Hirst, Di Cristo, & Espesser, 2000; Hirst, 2001), Uluslararası Sesbilim Abecesinin Fransızca tabanında ezgiye dayalı bir karşılığı olarak çokdilli bir çeviriyazı sistemi olarak oluşturulmuştur. Bu model çok başarılı olamamış olmasına karşın, sonraki modeller için dikkate değer fikirlerin üretimine zemin hazırlamıştır.

c. *Momel Ezgi Modeli* (Hirst & Espessser, 1993), günümüze kadar İngilizce, İspanyolca, Arapça, İtalyanca ve Fransızca gibi pek çok dilde çalışılmış ve Intsint Ezgi Modelinden yararlanılarak geliştirilmiştir. Bu algoritma F_0 eğrilerine dayalı olarak hedef noktalarını sıklık ve süre üzerinden incelemektedir.

d. *Fujisaki Üstdüşümsel Modeli* (Fujisaki & Kawai, 1982), ezgiyi öbek ve vurgu olmak üzere iki temel sistem üzerinden sentezlemektedir. Model, F_0 konturunu öbek ve vurgu kontrol mekanizmaları çerçevesinde ele alarak sade bir algoritma içermektedir. Ancak, sözlü üretim sırasında F_0 konturunda oluşabilecek olası kişiye özgü değişmezlere yönelik açıklama yapmakta yetersiz kalabilmektedir.

e. *Eğim (TILT) Modeli* (Taylor, 1992), ezgiyi olaylar dizisi olarak genlik ve süre ekseninde incelemektedir. Bu model ayrıca, F_0 perde örüntülerinin belirlenmesinde karmaşıklıktan uzak bir ezgi sentezleme düzeni sunmaktadır.

f. *PAintE Ezgi Modeli* (Möhler & Conkie, 1998; Möhler, 2001), s-biçiminde tanımlanan ezgi eğrilerini incelerken, vurgulu seslemleri referans noktaları olarak kabul eden uygulanabilirliği yüksek bir ezgi sentezleme modelidir.

g. Bu bölümün son kısmında ise, konuşma seslerinin yapay bir üretimini sunan *Metinden Konuşma Sentezleme* (Allen ve diğ., 1987) modelleri (bkz. Eklemeli Konuşma Sentezleme ve Birim Seçimi Tabanlı Konuşma Sentezleme (Hunt & Black, 1996); İstatistiksel Konuşma Sentezleme (Zen ve diğ., 2009); Hidden Markov Tabanlı Konuşma Sentezleme (Yoshimura ve diğ., 1999)) ana çizgileriyle betimlendikten sonra, Türkçe üzerine alanyazında günümüze kadar yapılmış araştırmalardan söz edilmiştir.

Son yıllarda yapay zeka araştırmalarının gelişimiyle birlikte, önem kazanan hesaplamalı sesbilgisi alanında ezgi sentezleme tabanında geliştirilen modellerin artışı günümüz araştırmalarına yadsınamaz bir katkı sağlamaktadır. Bilgisayar ve makine arakesiti araştırmalarına da katkı sunan bu modeller,

yazılı metinlerin dijital ortamda sözlü üretime dönüştürülmesinü inceleyen metinden konuşma sentezlemesi gibi pek çok farklı alanlar kapsamında ele alınarak, hesaplamalı sesbilgisi alanının gelişimine önem kazandırmaktadır.

Kaynakça

Abdullahbeşe, E. (2011). *Fundamental frequency contour synthesis for Turkish text-to-speech*. [Yayımlanmamış Yüksek Lisans Tezi]. Boğaziçi Üniversitesi.

Allen, J., Hunnicut, S., & Klatt, D. H. (1987). *From text to speech: The mitalk system*. Cambridge University Press.

Beckman, M. (1996). The parsing of prosody. *Language and Cognitive Processes (11),* 17-67.

Boersma P., & Weenink, D. (2020). Praat: Doing phonetics by computer [bilgisayar yazılımı]. Version 6.1.09. Retrieved from http://www.praat.org/.

Campione E., & Véronis, J. (1998). *A multilingual prosodic database*. Proceedings of ICSLP98, Sidney.

Campione, E., Flachaire, E., Hirst, D. J., & Véronis, J. (1997). *Stylization and symbolic coding of F0, a quantitative approach*. Proceedings of ESCA tutorial and research workshop on intonation. Atina.

Chiba, T., & Kajiyama, M. (1941). *The vowel, its nature and structure*. Tokyo-Kaiseikan Yayınevi.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge Universtity Press.

Crystal, D. (1992). *A dictionary of linguistics and phonetics*. (5. Baskı). Blackwell Publishers.

d'Alessandro, C. (2012). Voice source parameters and and prosodic analysis. İçinde S. Sudhoff, Lenertová, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, Richter, N., & Schlieβer, J. (Yay.), *Method in empirical prosody research (3)* (ss. 63-87). Walter de Gruyter.

Dalva, D. (2012). *Automatic speech recognition system for Turkish spoken language*. [Yayımlanmamış Yüksek Lisans Tezi]. Işık Üniversitesi.

Dalva, D., Revidi, İ. D., Güz, Ü., & Gürkan, H. (2014). *Türkçe haber yayını verileri için bürünsel bilginin çıkarılması ve cümle bölütlemede kullanıması*. 22th Signal processing and communications applications conference (SIU).

Ergenç, İ. (1989). *Türkiye Türkçesinin görevsel sesbilimi.* Engin Yayınları.

Ergenç, İ., & Bekar Uzun, İ. P. (2017). *Türkçenin ses dizgesi*. Seçkin Yayıncılık.

Friederici, A., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language* 89, 267-276.

Fujisaki, H., & Kawai, H. (1982). *Modeling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation*. Preprints of the working group on intonation, 13th international congress of linguists. (ss. 57-70). Tokyo.

Fujisaki, H., Ljungqvist, M., & Murata, H. (1993). *Analysis and modeling of word accent and sentence intonation in Swedish.* Computer science IEEE international conference on acoustics, speech, and signal processing.

Gessinger, I., Schweitzer, A., Andreeva, B., Raveh, E., Möbius, B., & Steiner, I. (2018). *Convergence of pitch accents in a shadowing task*. 9th International conference on speech prosody. 13-16 Haziran 2018. Polonya.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis, 2*(1), 23-68.

Güneş, G. (2015). *Deriving prosodic structures* [LOT Dissertation Series, 397], [Yayınlanmış Doktora Tezi]. LOT.

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. Mouton.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. The University of Chicago Press.

Hirst, D. J. (2001). Automatic analysis of prosody for multilingual speech corpora. İçinde E. Keller, G. Bailly, A. Monaghan, J. Terken, & M. Huckvale (Yay.), *Improvements in Speech Synthesis* (ss. 320-327). Wiley.

Hirst, D. J., & Di Cristo, A. (1998). *Intonation systems: A survey of twenty languages*. Cambridge University Press.

Hirst, D. J., & Espesser, R. (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix-en-Provence* 15, 75-85.

Hirst, D. J., Di Cristo, A., & Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. İçinde M. Horne (Yay.), *Prosody: Theory and experiment* (ss. 51-87). Springer.

Hirst, D. J., Ide, N., & Véronis, J. (1994). *Coding fundamental frequency patterns for multilingual synthesis with INTSINT in the MULTEXT Project*. Proceedings of 2nd ESCA workshop on speech synthesis. New York.

Hirst, D. J. (2004). *Lexical and non-lexical tone and prosodic typology*. Proceedings of international symposium on tonal aspects of languages: With emphasis on tone languages (ss. 81-88). Beijing, Mart 2004.

Hirst, D. J. (2007). A Praat plugin for Momel and Intsint with improved algorithms for modelling and coding intonation. ICPhS XVI. Saarbrücken, 6-10 Ağustos, 2007.

Hunt, A., & Black, A. (1996). *Unit selection in a concatenative speech synthesis system using a large speech database*. Proceedings of ICASSP (ss. 373-376).

İpek, C. (2015). *The phonology and phonetics of Turkish intonation*. [Yayımlanmamış Doktora Tezi]. University of Southern California.

Kamali, B. (2011). *Topics at the PF interface of Turkish*. [Yayımlanmamış Doktora Tezi]. Harvard Üniversitesi.

Kan, S. (2009). *Prosodic domains and the syntax-prosody mapping in Turkish*. [Yayımlanmamış Yüksek Lisans Tezi]. Boğaziçi Üniversitesi.

Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech*. (2. Baskı). Singular/Thomson Learning.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis and perception of voice quality variations among male and female talkers. *Journal of the Acoustical Society of America* 87*,* 820-856.

Külekçi, M. O., & Oflazer, K. (2006). An infrastructure for turkish prosody generation in text-to-speech. 15th Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN).

Ladd, D. R. (1996). *Intonational phonology*. Cambridge University Press.

Liberman, M. (1975). *The intonation system of English*. [Yayınlanmış Doktora Tezi, Indiana University]. Indiana University Press.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry 8*(2), 249-336.

Malkoç, E. (2011). *Konuşmanın kişiye özgü değişmezleri: ünlüler üzerine sesbilgisel bir çalışma*. [Yayımlanmamış Doktora Tezi]. Ankara Üniversitesi.

Möhler, G. (2001). Improvements of the PaIntE model for F0 parametrization. Technical Report, Institute for Natural Language Processing, Ms., University of Stuttgart.

Möhler, G., & Conkie, A. (1998). *Parametric modeling of intonation using vector quantization*. Proceedings of 3rd International ESCA/COCOSDA workshop (ETRW) on speech synthesis.

Navas, E., Hernaez, I., & Sanchez, J. M. (2000). *Subjective evaluation of synthetic intonation*. Proceedings of 2002 IEEE workshop on speech synthesis (2002). (ss. 23-26). Santa Monica.

O'Connor, J. D., & Arnold, G. F. (1973). *Intonation of colloquial English*. Longman.

Oskay, B. (2002). *Automatic modelling of Turkish prosody*. [Yayımlanmamış Yüksek Lisans Tezi]. Orta Doğu Teknik Üniversitesi.

Oskay, B., Salor, Ö., Özkan, Ö., Demirekler, M., & Çiloğlu, T. (2001). *Türkçe tümceler için metinden ezgi belirlenmesi ve uygulaması*. IEEE 9. Sinyal işleme ve iletişim uygulamaları kurultayı (SİU-2001), KKTC.

Öztürk, Ö. (2006). *Modeling phoneme durations and fundamental frequency contours in Turkish speech*. [Yayımlanmamış Doktora Tezi]. Orta Doğu Teknik Üniversitesi.

Palmer, H. (1922). *English intonation with systematic exercises*. Cambridge University Press.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. [Yayınlanmış Doktora Tezi]. MIT. Indiana University Linguistics Club Press.

Pierrehumbert, J., & Liberman, M. (1984). Intonational invariance under changes in pitch range and length. İçinde M. Aronoff, & Oehrle, R. *Language sound structure: Studies in phonology*. MIT Yayınları.

Rossi, P., Palmieri, F., & Cutugno, F. (2002). *A method for automatic extraction of Fujisaki-model parameters*. Proceedings of Speech Prosody (2002), Aix-en-Provence (ss. 615-618).

Salor, Ö. (2005). *Voice transformation and development of related speech analysis tools for Turkish*. [Yayımlanmamış Doktora Tezi]. Orta Doğu Teknik Üniversitesi.

Şayli, Ö. (2002). *Duration analysis and modeling for Turkish text-to-speech synthesis*. [Yayımlanmamış Yüksek Lisans Tezi]. Boğaziçi Üniversitesi.

Schweitzer, K., Walsh, M., & Schweitzer, A. (2017) To see or not to see: Interlocutor visibility and likeability influence convergence in intonation. *Interspeech* 2017, 919-923.

t'Hart J., & Cohen, A. (1973). Intonation by rule: A perceptual quest. *Journal of Phonetics, 1*(1973), 309-327.

Taylor, P. A. (1992). *A phonetic model of English intonation*. [Yayınlanmış Doktora Tezi, University of Edinburgh]. Indiana University Linguistics Club Press.

Taylor, P. A. (1995). The rise/fall/connection model of intonation. *Speech Communication* 15, 169-186.

Taylor, P. A. (2009). *Text-to-speech synthesis*. Cambridge University Press.

Türk, O. (2003). *New methods for voice conversion*. [Yayımlanmamış Yüksek Lisans Tezi]. Boğaziçi Üniversitesi.

Türk, O. (2007). *Cross-lingual voice conversion*. [Yayımlanmamış Doktora Tezi]. Boğaziçi Üniversitesi.

Uslu, İ. B., & İlk, G. (2012). *Türkçe metinden konuşma sentezlemede kural tabanlı duyumsal bir ezgi modeli*. 20th Signal processing and communications applications conference (SIU). Muğla.

Vural, E. (2003). *A prosodic Turkish text-to-speech synthesizer*. [Yayımlanmamış Yüksek Lisans Tezi]. Sabancı Üniversitesi.

Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., & Kitamura, T. (1999). *Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis*. Proceedings of eurospeech conference (ss. 2347-2350).

Zen, H., Tokuda, K., & Black, A. (2009). Statistical parametric speech synthesis. *Speech Communication, 51*(11), 1039-1064.

Kısaltmalar

F_0	Temel Sıklık
(%)	Sınır İşaretli Ton
(-)	Öbek İşaretli Ton
(*)	Yıldız İşaretli Ton
MKS	Metinden Konuşma Sentezleme
ToBI	(*Tones and Break Indices*)

Bölümde Kullanılan Terimler Listesi

Açılım	Antinode
Akustik Sesbilgisi	Acoustic Phonetics
Alçak	Fall / Low
Anahtar	Key
Bağlantı Noktası	Connections
Baş	Head
Baş-Öncesi	Pre-Head
Bilgisayarbilimleri	Computer Sciences
Birim Seçimi Tabanlı Konuşma Sentezleme	Unit Selection Speech Synthesis
Boğum	Node
Bürün	Prosody
Bürün Modellemesi	Prosodic Modelling
Bürün Sentezlemesi	Prosody Synthesis
Çekirdek Vurgu	Nuclear Accent
Daralma Kuramı	Perturbation Theory
Durak	Pause
Düzen	Range
Düzeye Dayalı Tonlama	Downdrift
Eğim	Declination / Tilt
Eğim Modeli	Tılt Model
Eklemeli Konuşma Sentezleme	Concatenative Approach to Text-to-Speech Synthesis
Eklemleyici Sesbilgisi	Articulatory Phonetics
Eklenti	Tail
Eklenti Komut Dizisi	Plugin

Enerji Yoğunluğu	Intensity
Ezgi	Intonation
Ezgi Öbeği	Intonational Phrase
Ezgisel Sesbilim	Intonational Phonology
F_0	Fundamental Frequency/F_0
F_0 Değerine Erişim için Öbek-Komut Girdisi	Phrase-Command Contribute to F_0
F_0 Değerine Erişim için Vurgu-Komut Girdisi	Accent-Command Contribute to F_0
F_0 Konturlarının Sadeleştirmesi	Stylization
FIR Süzgeci	Finite Impulse Filters
Fujisaki Üstdüşümsel Modeli	Fujisaki's Superpositional Model
Genlik	Amplitude
Gerginlik-Gevşeklik	Tense-Lax
Gırtlak Formantları	Glottal Formants
Hedef Noktaları	Target Points
Hesaplamalı Dilbilim	Computational Linguistics
Hesaplamalı Sesbilgisi	Computational Phonetics
Hidden Markov Tabanlı Konuşma Sentezleme	Hidden Markov Model based Text-To-Speech Synthesis
Hollanda Ezgi Okulu	Dutch School Of Intonation
İkili Ses	Diphone
İngiliz Ezgi Okulu	British School Of Intonation
INTSINT Ezgi Modeli	**In**ternational **T**ranscription **S**ystem for **Int**onation
İşitsel Sesbilgisi	Auditory Phonetics
İstatistiksel Konuşma Sentezleme	Statistical Parametric Speech Synthesis
Kontur	Contour
Konuşma Seslerinin Sentezlenmesi	Speech Synthesis
Konuşmacı Tanıma	Speech Recognition
Konuşmacı Tanımlama	Speech Recognition
Metinden Konuşma Sentezleme	Text-to-Speech Synthesis
MOMEL Ezgi Modeli	Modélisation De Melodie
Öbek	Phrase
Öbek İşaretli Ton	Phrase Tones
Öbek-Komut Girdisi	Phrase-Command İnput

Öbeksel Belirginlik	Phrasal Prominence
Odak	Focus
Olay Dizileri	Event Series
Ölçülü Kuram	Metrical Theory
Özerkparça Kuramı	Autosegmental Theory
Özerkparçalı-Ölçülü Sesbilim	Autosegmental-Metrical Phonology
PAintE Ezgi Modeli	**Pa**rametric Representations of **Int**onational **E**vents
Perde Vurgusu	Pitch Accent
Saf Alçaklık	Pure Fall
Saf Yükseklik	Pure Rise
S-Biçimi	Sigmoid
Ses Dönüştürme	Voice Conversion
Ses Rengi	Timbre
Ses Yükseklik Düzeyi	Pitch Range
Sesbilgisi	Phonetics
Sınır	Boundary
Sınır İşaretli Ton (%)	Boundary Tones
Söylem	Discourse
Spektral Eğim	Spectral Tilt
Spektral Eğim	Slope
Spektral Tepe Noktaları	Spectral Peaks
Süre	Duration
Temel Sıklık	Fundamental Frequency
Ton	Tone
Türkçenin Bürün Dizgesinin Özdevinimli Modellemesi	Automatic Modelling of Turkish Prosody
Uluslararası Sesbilim Abecesi	International Phonetic Alphabet
Vurgu	Accent
Vurgu-Komut Girdisi	Accent-Command Input
Yukarıdan Aşağıya İniş	Downstep
Yüksek	Rise / High
Yükselen/Alçalan/Bağlantı Modeli	Rise/Fall/Connection Model