

Görme Engelliler için Türkçe Metinden Konuşma Sentezleme Yazılımı Geliştirilmesi

Benian TEKİNDAL¹, Güray ARIK²

¹ Bilgisayar Eğitimi Bölümü, Gazi Üniversitesi, Ankara, Türkiye

² Bilgisayar Eğitimi Bölümü, Gazi Üniversitesi, Ankara, Türkiye

beniantekindal@gmail.com, gurayarik@hotmail.com

(Geliş/Received: 13.06.2011; Kabul/Accepted: 16.10.2012)

Özet— Görme engelliler, bilgisayarı kullanmak için engellerine yönelik geliştirilmiş destek teknolojilere ihtiyaç duyarlar ve en büyük yardımcıları konuşma sentezleme yazılımlarıdır. Günümüzde, daha çok batı dillerine yönelik dünya çapında geliştirilmiş sözcük tabanlı onlarca konuşma sentezleme yazılımı mevcuttur. Ancak Türkçenin batı dillerinden farklı olarak sona eklemeli bir dil olması ve yazıldığı gibi okunan bir dil olması sebebiyle bu yazılımlarda dil desteği bulunmamakta veya sentezlenen Türkçe metinler anlaşılır olmamaktadır. Bu nedenle Türkçe, yapısı itibariyle farklı yöntemlerle çalışılması gereken bir dildir. Konuşma sentezleme teknikleri incelendiğinde, hecenin ses birimi olarak kullanılacak bir yapı Türkçe için oldukça uygun olacaktır. Çalışmanın temel amacı, görme engellilerin bilgisayar kullanmalarını etkinleştirmektir. Bu amaca yönelik olarak, daha kolay ve daha pratik kesimleme ve etiketleme yapısı sunan, parça tabanlı ardışık eklemeli(concatenative) konuşma sentezleme tekniği kullanarak hece tabanlı Türkçe metinden konuşma sentezleme yapabilen Okuyucu TTS isimli bir metin okuyucu yazılımı geliştirilmiştir. Yazılım Visual C#.NET programlama dili ile kodlanmıştır. Geliştirme ortamı olarak Visual Studio 2008 Professional kullanılmıştır. Yazılım tamamlandıktan sonra görme engellilerin ihtiyaçlarına cevap verip vermediğine yönelik doğrulama ve geçerleme çalışmaları 10 görme engelli kişinin katılımı ile yapılmıştır. Yapılan doğrulama ve geçerleme çalışmaları sonucunda yazılımda üretilen konuşmanın kalitesi MOS benzeri bir derecelendirmeden 5 üzerinden 3,7 puan almıştır.

Anahtar Kelimeler— Konuşma sentezleme, görme engelli, Türkçe heceleme sistemi

Development of Speech Synthesis Software From Turkish Text for the Visually Impaired

Abstract— Visually impaired, need to use the computer technologies, enhanced support for their disability and the most important aids are speech synthesis software. Today, throughout the world particularly for Western languages, dozens of enhanced word-based speech synthesis software is available. These softwares do not have language support or synthesized Turkish texts is not clear due to the Turkish is an agglutinative language and different from western languages, since it is a language that is read as it is written. Though, Turkish is a language that has to be studied in different ways due to the structure. Syllable structure used as a unit of volume is considered to be highly suitable for Turkish, when speech synthesis techniques are examined. This study aims to enable the visually impaired to use computers effectively. For this purpose, a text reader software called Okuyucu TTS was developed, providing an easier and more practical structure segmentation and labeling, using track-based sequential insertion (concatenative) speech synthesis technique and capable syllable-based speech synthesis from Turkish text. Software was coded with Visual C#.NET programming language. Visual Studio 2008 Professional was used as a development environment. After the software developed, to measure whether the needs of the visually impaired, verification and validation studies conducted with the participation of 10 visually impaired people. As a result of the verification and validation studies, the quality of speech produced ranked 3,7 out of 5 in a MOS-like rating system.

Keywords— Turkish hyphenation algorithm, speech synthesis, visually impaired

1. GİRİŞ

Günümüzde internetin büyük bir bilgi havuzuna ve sosyal paylaşım platformuna dönüşmesiyle birlikte,

bilgisayarın günlük yaşamda kullanımı hızla artmaktadır. Bu durum insanların günlük yaşamlarındaki kitap okuma, iletişim kurma, ders

çalışma gibi biçimsel alışkanlıkları ile sosyal yaşantılarını da değiştirmektedir. Bu değişim normal bireyler ve ortopedik engelli gibi bazı engelli bireyler için çok büyük avantajlar getirmekle birlikte; görme engelli bireyler için, özellikle bilgiye erişme konusunda, büyük dezavantajları beraberinde getirmektedir. Çünkü görme engelliler bilgisayarı normal bireyler gibi kullanamazlar. Bilgisayarı kullanmak için engellerine yönelik geliştirilmiş bazı destek teknolojilere ihtiyaç duyarlar. Bu destek teknolojiler içerisinde, görme engellilerin en büyük yardımcıları ekran okuyucu programlardır. Bu programlar, bilgisayar ekranında görüntülenen metinleri doğal insan sesi kalitesiyle seslendirerek, görme engelli bireyi yönlendirir ve bilgisayar teknolojisinin sunduğu imkânlardan faydalanmasını sağlar. Ekran okuyucu programları aslında bir konuşma sentezleme yazılımıdır[1]. Konuşma sentezleme ise bilgisayarlarda kullanılan metni konuşmaya çeviren sistemlere verilen isimdir[2].

Günümüzde daha çok batı dillerine yönelik dünya çapında geliştirilmiş sözcük tabanlı onlarca konuşma sentezleme yazılımı mevcuttur. Ancak Türkçenin Latin kökenli batı dillerinden farklı olarak sona eklemeli bir dil olması[3] sebebiyle çoğu yazılımda dil desteği bulunmamaktadır. Bununla birlikte, az sayıda istisnası olmakla birlikte, Türkçenin yazıldığı gibi okunan bir dil olması sebebiyle bu yazılımlarda sentezlenen Türkçe metinler anlaşılır olmamaktadır. Bu nedenle Türkçe, yapısı itibarıyla farklı yöntemlerle çalışılması gereken bir dildir.

Dünya çapında üç farklı metinden konuşma sentezleme tekniği bulunmaktadır. Bu teknikler; articulatory (boğumlama) sentezleme, format (biçimsel) sentezleme ve concatenative (eklemeli) sentezlemedir. Bu teknikler arasında, Türkçenin sona eklemeli ve hece tabanlı bir dil yapısı olması sebebiyle concatenative (eklemeli) sentezleme tekniği ile hecenin ses birimi olarak kullanılacağı bir yapı, Türkçe metinden konuşma sentezleme için oldukça uygun olacaktır.

Eklemeli(concatenative) sentezleme tekniğinde konuşma sentezi, ses veri tabanına daha öncesinde seslendirilerek saklanmış ses kayıt parçalarının birbirine ardışık olarak eklenmesi ile yapılır. Konuşma sentezi yapılacak girdi metni analiz edilerek, ardışık olarak eklenecek ses parçaları belirlenir. Bu tarz bir sentezleme sürecindeki en önemli kısım, ses veri tabanının oluşturulması ve kesimlemedir. Ses veri tabanını oluşturacak ses kayıtları, diksiyonu düzgün bir konuşmacıya seslendirme yaptırılarak oluşturulur. Oluşturulan ses kayıtları, geliştirilen sistemin kurallarına göre kesimlenerek ses veri tabanında etiketlenir. Kesimleme yapılacak ses birimi, sözcük veya hece tabanlı olabilir.

Türkçe dili için sözcükleri ses birimi olarak kullanmak mümkün görünmemektedir. Çünkü Türkçe sona eklemeli bir dildir ve bir sözcükten ek getirmek suretiyle aynı veya farklı anlamda birçok sözcük

türetilebilir[4]. Şu an için, Türkiye Türkçesi'nin en gelişmiş sözlüğü Büyük Türkçe sözlükte söz, deyim, terim ve isim olmak üzere toplam 616 767 sözcük bulunmaktadır[5]. Bu sözcüklerin sona alacağı eklerin de olacağı düşünülürse, bu durumda on milyonlarca kelimenin kesimlenerek ses veri veritabanında saklanması gerekecektir. Ayrıca ses veri tabanı programın içerisine gömülü olacağından, sözcük tabanlı yapılacak kesimleme, geliştirilecek programın boyutunu çok ciddi oranda artıracaktır. Hece tabanlı yapılacak bir kesimlemede ise, dildeki tüm hece alternatifleri ses kaydının içerisinden kesilerek elde edilebilir. Sözcük tabanlı kesimlemeye göre daha pratik bir yöntem olmakla birlikte; Türkçede beş harften oluşan hece olduğu düşünülürse (kramp gibi), bu kesimleme ve etiketleme yöntemi de uzun bir süreç gerektirecektir. Bu nedenle ses veri tabanının boyutunun düşürülmesi ve kesimleme süresinin azaltılmasına yönelik daha pratik yöntemler araştırılmalıdır.

2. YÖNTEM VE MATERYAL

Konuşma sentezleme sistemlerinin günümüzde çok yaygın kullanım alanları vardır ve gün geçtikçe kullanım alanları artmaktadır. Konuşma sentezleme yazılımları, görme engellilere elektronik ortamda bulunan bir kitabı, dergiyi, internet sayfasını insan sesine yakın bir şekilde okuyacak şekilde tasarlanabilir. Bu çalışmada, konuşma sentezleme teknolojileri kapsamında, özellikle ses veri tabanı boyutunu azaltarak kesimleme ve etiketleme maliyetini ciddi oranda düşüren Türkçe dil yapısına uyumlu daha farklı ve daha pratik bir kesimleme ve etiketleme yöntemi üzerinde durulmaktadır. Bu doğrultuda, görme engellilere yönelik olarak Türkçe konuşma sentezleme yapabilen hece tabanlı bir metinden konuşma sentezleme yazılımı geliştirilmiştir. Geliştirilen yazılımın temel amacı, görme engellilerin kitap okuma, sınavlara hazırlık gibi eğitim ihtiyaçları kapsamında bilgisayar ortamında bulunan çeşitli formatlardaki elektronik kitapları daha öncesinde seslendirmeye ihtiyaç duymadan insan sesine yakın bir şekilde konuşmaya çevirmektir.

Geliştirilen yazılım Visual C#.NET programlama dili ile kodlanmıştır. Geliştirme ortamı olarak Visual Studio 2008 Professional kullanılmıştır. Konuşma sentezi, ses veri tabanında saklanan ses kayıt parçalarının birbirine eklenmesiyle yapılır. Sentezlemesi yapılacak girdi metni, bir metin işleme sürecinin ardından ses veri tabanına karşılık gelen hecelere ayrılmaktadır.

Hece tabanlı metinden konuşma sentezleme yazılımının geliştirilmesi sürecinde, Türkçe dil yapısı da detaylı olarak incelenmiştir. Hecenin ses birimi olarak kullanılacağı bir yapının Türkçe için oldukça uygun olacağı değerlendirilmiştir. Türkçe dil yapısı incelendiğinde, bir girdi metnin kelimelere ayrıldıktan sonra her bir kelimenin otomatik olarak hecelerine ayrılacağı anlaşılmıştır. Türkçe

heceleme sistemi basit ve mekanik bir yapıya sahiptir ve mevcut kurallara uyularak tasarlanacak bir algoritma ile ifade edilebilir. Bu maksatla kelimeleri Türkçe dilbilgisi kurallarına uygun olarak heceleylebilen ve yazılım içine entegre edilebilen bir “Türkçe Heceleme Algoritması” geliştirilmiştir. Geliştirilen algoritma yardımıyla hecelere ayrılan girdi metni, daha öncesinde seslendirilen ve ses veri tabanında etiketlenerek saklanan ses kayıtları ile eşleştirilerek sentezlenmektedir. Bu nedenle geliştirilen heceleme algoritmasının, kelimeleri hecelerine dil bilgisi kurallarına uygun olacak şekilde ayırması önem arz etmektedir. Bunu sağlamak için Türkçe hece yapısı detaylı olarak incelenmiştir.

2.1. Türkçe Hece Yapısı

Ağzın bir hareketiyle bir defada söyleneblen sözcük parçalarına hece denir. Heceler harflerden oluşur. Türkçede hece sistemi genel olarak mekanik bir yapıya sahiptir. Bu mekanik yapı belirli kurallarla donatılmıştır. Ana kurallar şu şekilde sıralanabilir [6].

- Türkçedeki en kısa hece bir sesli harften oluşmaktadır. Örneğin “O” kelimesi kendi başına bir hece oluşturur.
- Birden çok sessiz harften hece olmaz. İçinde mutlaka bir sesli harf olmalıdır. Batı kökenli kelimeler istisna oluşturur.
- Türkçe kelimelerde en fazla iki adet sessiz harf yan yana gelebilir. Batı kökenli kelimeler istisna oluşturur.
- İki sessiz harf yan yana geldiğinde ve hiç sesli harf olmadığında iki ayrı hece oluşur.
- Türkçede kelime içinde iki sesli arasındaki sessiz, kendinden sonraki sesliyle hece kurar: a-ra-ba, bi-çi-mi-ne, in-sa-nın, ka-ra-ca, alt-lık, al-dı.
- Kelime içinde yan yana gelen sessizlerden sonuncusu kendisinden sonraki sesliyle, diğerleri kendilerinden önceki sesliyle hece kurar: bir-lik, sev-mek, Türk-çe, Kork-maz.
- Batı kökenli kelimeler, Türkçenin hece yapısına göre hecelere ayrılır: band-rol, kont-rol, prog-ram, sant-ral, sürp-riz, tund-ra, volf-ram.

Bu ana kurallar doğrultusunda en uzun ve en kısa heceler Tablo 1’de görülebilir. Sembolik olarak sesli harfler “a” ile sessiz harfler “b” ile belirtilmiştir. Tablo 1’de görüldüğü gibi Türkçe dilinde hece sisteminin 9 farklı hece tipinde olabileceği anlaşılmaktadır.

Tablo 1. Türkçe hece tipi yapıları

S.Nu.	Hece Sembolü	Örnek
1	a	o
2	ab	iş, ev
3	ba	su, bu
4	abb	alt, aşk
5	bab	kış, sel
6	bba	gri
7	babb	kalp, sarp
8	bbab	krem, tren
9	bbabb	kramp, branş

Belirtilen hece sistemi kuralları dilimizde bulunan yabancı kökenli kelimeler nedeniyle bazen sekteye uğrayabilmektedir. Sayıları az olmakla birlikte özellikle üç sessiz harfin yan yana geldiği heceler barındıran kelimeler bulunmaktadır. “Stra-tos-fer”, “stres”, “ekst-re”, “me-ga-hertz” gibi yabancı kökenli kelimeler bu kuralı bozan kelimelere örnek olarak verilebilir.

Hece tipi yapıları bu kurallar doğrultusunda incelendiğinde Türkçede anlamlı veya anlamsız oluşturulabilecek hece sayısı 1 714 952’dir. Tablo 2’de tespit yöntemi verilmiştir. Mevcut kurallar doğrultusunda tespit edilen anlamlı veya anlamsız bu kadar hecenin seslendirilerek ve etiketlenilerek kaydedilmesi büyük çabalar gerektirmektedir.

Tablo 2. Toplam hece sayısı

S.Nu.	Hece Sembolü	Örnek	Çarpan	Toplam
1	a	o	8	8
2	ab	iş, ev	8x21	168
3	ba	su, bu	21x8	168
4	abb	alt, aşk	8x21x21	3 528
5	bab	kış, sel	21x8x21	3 528
6	bba	gri	21x21x8	3 528
7	babb	kalp, sarp	21x8x21x21	74 088
8	bbab	krem, tren	21x21x8x21	74 088
9	bbabb	kramp, branş	21x21x8x21x21	1 555 848
GENEL TOPLAM				1 714 952

2.2. Etiketleme ve Ses Veri Tabanının Oluşturulması

Etiketleme, elde edilen ses birimlerinin veri tabanında isimlendirilerek saklanmasıdır. Bu çalışma kapsamında geliştirilen konuşma sentezleme yazılımında Türkçe dil yapısına uyumlu ve daha farklı, daha pratik bir kesimleme ve etiketleme yöntemi sunulmaktadır. Tüm hece kombinasyonlarının ses kaydından kesimlenmesi yerine sadece bir ve iki harfli hece kombinasyonları seslendirilerek hecenin başlangıç ve bitiş sınırları belirlenir ve ses veri tabanına hece ismiyle etiketlenir. Yazılımın çalışma anında, ikiden büyük harfli heceler önce ikili hece formatına dönüştürülerek uygun ikili hece etiketi belirlenir, ardından fazladan harf kaldırılarak uygun ses üretilir.

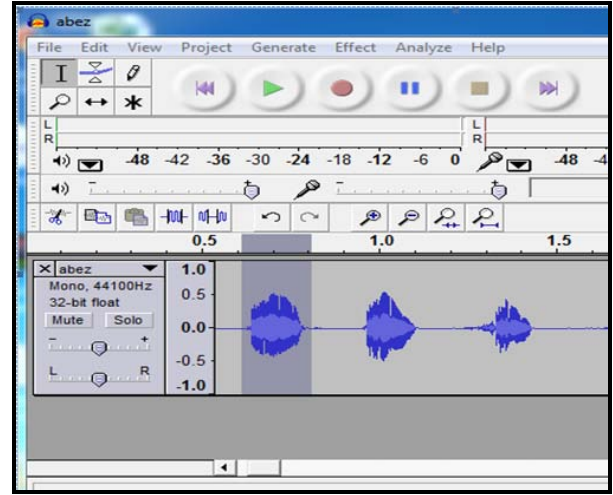
Örneğin üç harften oluşan “ZAR” hecesi ses veri tabanında etiketlenmemiş bir hecedir. Ancak ses veri tabanında “ZA” ve “AR” olarak etiketlenmiş iki harften oluşan iki ayrı hece bulunmaktadır. “ZAR” hecesi, seslendirilebilmek için yazılımın çalışma anında ilk olarak “ZA” ve “AR” olarak ikili hece formatına dönüştürülür. İlk oluşan “ZA” hecesi, ses veri tabanında kendisine karşılık gelen ses ile eşleştirilir. “AR” olarak üretilen iki harfli ikinci hece ise sözde hece olarak işaretlenir. Ses veri tabanı oluşturulması aşamasında, normal hece olarak seslendirilen 21 adet sesiz harfin, ayrıca ikili hecelerde son harf olarak seslendirilmiş bölümü kesimlenerek etiketlenmiştir. Böylece sözde hece olarak işaretlenen “AR” hecesindeki fazladan üretilen “A” harfi atılarak “ZA-R” olarak seslendirilebilmektedir. Eşleştirme yapılımasını sağlamak için yazılımsal olarak fazladan üretilen sesli harfin atılması yine yazılımsal olarak yapılmaktadır. Bu yöntemle herhangi bir sıkıştırma algoritması kullanmadan yaklaşık 5 GB’lık yer kaplayacak olan ses veri tabanı 10 MB’a inmiştir. Bunun yanında kesimleme ve etiketleme süreci de zaman açısından çok ciddi oranda azaltılmaktadır. Bu kapsamda ses veri tabanına kaydedilmesi gereken heceler Tablo 3’te görülmektedir. Tabloda “a” harfi sesli harfi, “b” harfi ise sessiz harfi temsil eder.

Tablo 3. Etiketlenmiş sesli hece şablonları

S.Nu.	Hece Sembolü	Örnek	Çarpan	Toplam
1	a	O	8	8
2	ab	iş, ev	8x21	168
3	ba	su, bu	21x8	168
4	ba	z dağılımı	21	21
5	-b	va-r	21	21
GENEL TOPLAM				386

Konuşma organlarının, anlamlı sözcükler ve tümceler oluşturmak için düzenli olarak çalışması sonucu çıkan birimlere sesbirim veya fon denir. Fon bir başka

ifadeyle dildeki anlam ayırıcı en küçük öge olarak da adlandırılabilir. Her dilin kendine özgü bir fon tanımı vardır[7]. Ortak olan ayırıcı olma özelliğidir. Türkçe fonem tabanlı bir dildir. Bu sebeple Türkçe için her harfin bir foneme karşılık geldiği söylenebilir. Bu çalışma kapsamında el ile kesimleme yapılırken fonem düzeyinde kesimleme yapılmıştır. Bunun için ses ifadelerini kaydetmek ve düzenlemek için açık kaynak kodlu bir yazılım olan Audacity programı kullanılmıştır. Programın bir görünümü Şekil 1’de görülmektedir.



Şekil 1. Audacity programında kayıt düzenleme

Orta alandaki bölüm, kesimlemesi yapılacak ses dalgasını gösterir. Bu bölümde sınırları belirlenmiş fonemler(heceler) kesimlenerek kendi ismiyle etiketlenir. Bu etiketleme ile oluşan ses dosyaları “wav” formatında olacak şekilde kayıt altına alınır. Örneğin; “ya” hecesi “ya.wav” isimli etiketle ses veri tabanına kaydedilir. Yapılan tüm bu işlemler ses veri tabanının oluşturulması ve etiketleme sürecini oluşturmaktadır.

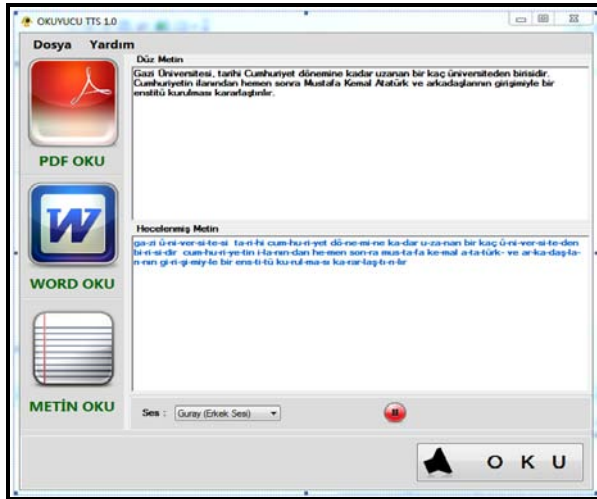
3. METİNDEN KONUŞMA SENTEZLEME YAZILIMI

Görme engelli bireyler hızla gelişen teknolojiden payını alamamaktadır. Çünkü gelişen teknoloji büyük oranda görseldir. Bu gelişen teknoloji içerisinde, görme engellilerin en önemli sıkıntılardan birisi eğitim ihtiyacı içerisinde değerlendirilen kitap okuma imkânlarının azalmasıdır. Günümüzde teknolojinin gelişmesi ve donanımın küçülmeye başlaması ile elektronik kitap sayısı normal basılı kitap sayısını geçmek üzeridir. Dünyanın en büyük internet üzerinden kitap satışı yapan şirketi olan Amazon, 2010 yılının ilk üç ayında satılan elektronik kitap sayısının matbu kitap sayısını geçtiğini açıklamıştır.

Görme engelli bireyin kitap okuyabilmesi için kitabın kabartma yazı olarak ifade edilen Braille alfabetiyle basılmış olması gerekmekte, ya da bir yardımseverin kitabı görme engelliye okuması gerekmektedir. Günümüzde görme engelli çocuklar için öykü ve masallar, ilköğretim ve lise çağındaki görme engelliler

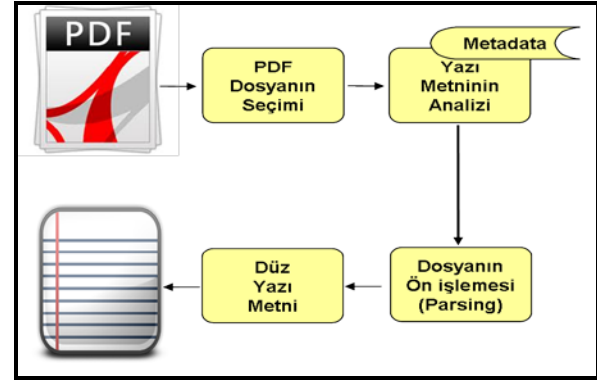
için hikâye ve romanlar ile özellikle sınava hazırlık dokümanları, üniversite ve yüksek lisans seviyesindeki görme engelliler için ise bölümlerine yönelik teknik sesli kitaplara ihtiyaç vardır. Kitap seslendirme işleminin sabır isteyen bir süreç olmasının yanında, bu geniş yelpazede ve özellikle eğitim sistemindeki yerleştirme sınavlarının müfredatlarının sıklıkla değiştiği günümüzde kitapların güncelliğinin sağlanması sürekli bir yapılanma gerektirmektedir.

Yukarıdaki bahsedilen sorunlar doğrultusunda; görme engellilerin kitap okuma, sınavlara hazırlık gibi eğitim ihtiyaçları kapsamında, bilgisayar ortamında bulunan çeşitli formatlardaki elektronik kitapları okuyabilen "Okuyucu TTS" isimli bir hece tabanlı metinden konuşma sentezleme yazılımı geliştirilmiştir. Geliştirilen yazılım elektronik ortamda bulunan bir dokümanı daha öncesinde seslendirmeye ihtiyaç duymadan, insan sesine yakın bir ses vererek konuşmaya çevirebilmektedir. Şekil 2'de, geliştirilen yazılım için tasarlanan kullanıcı ara yüzü görülmektedir.



Şekil 2. Okuyucu TTS 1.0 yazılımı kullanıcı ara yüzü

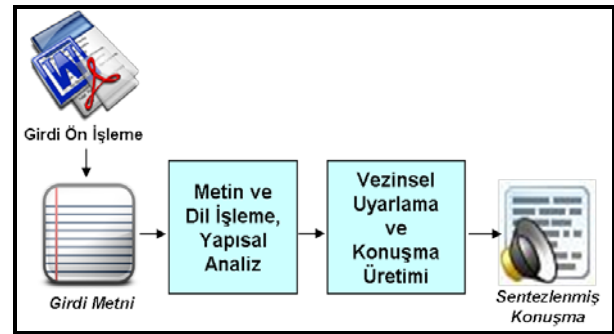
Elektronik kitap okuma işlemi tek bir elektronik dosya formatıyla kısıtlı değildir. Örneğin pdf, doc, docx ve txt tipindeki farklı formattaki elektronik dosyaları da sentezleyebilmektedir. Bu sentezleme süreci, girdi metninin ön işleme alınmasından sentezlenmiş konuşmanın çıkışına kadar kendi içerisinde bazı alt süreçleri barındıran bir yapıdır. Geliştirilen yazılımda; girdi metninin işlenmesi ve ses verilerinin çalışma anında eklenerek okuma çıktısı oluşturulması süreci tüm elektronik dosya formatları için sabittir. Ancak her elektronik dosya formatı farklı bir metadetaya sahiptir. Bu nedenle önemli olan husus girdi metninin, bir ön işlemeye tabi tutulmasıdır. Seçilen elektronik dosya formatı standart bir yapıya dönüştürülmelidir. Kullanıcı ara yüzünde yapılacak elektronik dosya formatının seçimini müteakip, yazılımın çalışma anında seçilen elektronik dosya tipine göre bir girdi ön işleme süreci yaratılır ve seçilen dosya standart bir yapıya dönüştürülür. Bir PDF uzantılı dosyanın girdi ön işleme süreci Şekil 4'te grafiksel olarak açıklanmıştır.



Şekil 4. PDF uzantılı dosyanın girdi ön işleme süreci

3.1 Konuşma Sentezleme Süreci

Konuşma sentezleme süreci temelde iki ana aşamaya ayrılabilir. İlk aşama metin ve dil işleme sürecidir. Girdi metni, dilin özelliğine göre fonetik bir yapı içerisinde uyarlanarak işlenir. Bu uyarlama işlemi sözcük tabanlı veya fonem tabanlı olabilir. İkinci aşama ise konuşmanın üretilmesi aşamasıdır. Bu akustik çıktı metnin analizi aşamasında oluşturulan fonetikten ve vezinsel uyarlama ön analizinden üretilir. Şekil 3'te bu sürecin basitleştirilmiş bir şekli verilmektedir.

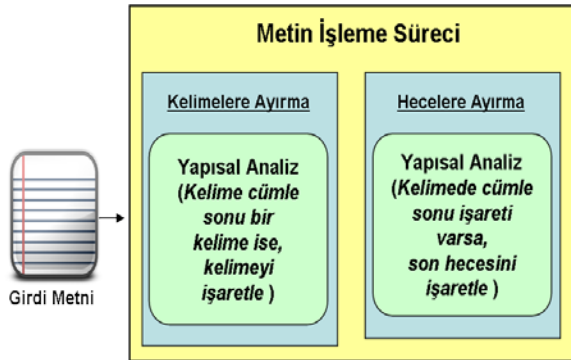


Şekil 3. Konuşma sentezleme süreci

Girdi dosyası, girdi ön işleme sürecine girerek standart bir girdi metnine dönüştürüldükten sonra, metin ve dil işleme sürecine gelir. Bu süreç dilin yapısına göre şekillenir. Türkçe sona eklemeli bir dildir ve kendine özgü bir metin işleme süreci gerçekleştirilmelidir. Metin işleme sürecinde yapılacak işlemler öncelikle metni noktalama işaretleri yardımıyla kelimelere ayırma, elde edilen kelimeleri de hecelere ayırma süreçlerini kapsamaktadır. Bu süreçte en önemli adım metnin yapısal analizinin yapılmasıdır.

Girdi metninin yapısal analizindeki amaç, girdi metninin ses dosyası haline getirildiğinde kısmen de olsa ezgisel bir konuşma olmasını sağlamaktır. İnsan konuşurken ya da bir metni okurken konuşmanın sonuna kadar nefesini tutmaz. Belirli aralıklarda duraksayarak nefesini ayarlamak zorundadır. Yazılı bir metinde bu duraksamaları yapmak ve okumanın anlaşılmasını sağlamak için noktalama işaretleri kullanılır[8].

Yapısal analiz süreci, metin işleme sürecinin bir alt sürecidir. Metin kelimelere ve hecelere ayrılma aşamasında; nokta, ünlem, soru işareti, iki nokta üst üste, üç nokta gibi noktalama işareti ile karşılaşılırsa önceki kelime cümle sonu kelimesi, kelimedeki son hece de cümle sonu hecesi olarak özel bir karakterle isimlendirilir. Okuma süresinde hecelerin birleştirilmesi aşamasında bu özel karakter ile cümle sonu kelimeler ve kelime sonu heceler tanınmakta ve buna göre konuşmaya duraksamalar verilmektedir. Virgül, noktalı virgül gibi diğer noktalama işaretleri içinde benzer bir özel işaret tanımlanmakta ve buna göre duraksama süresi ayarlanmaktadır. Girdi metninin yapısal analizi Şekil 5'te gösterilmektedir.



Şekil 5. Girdi metninin yapısal analizi

Metnin yapısal analizi yazılımsal olarak ayrı bir süreç olarak tanımlanabilir. Ancak metin işleme süreci içinde yapıldığında performans açısından daha verimli olur. Yapısal analiz sürecinde noktalama işaretlerinin tespiti için tüm girdi metninin gezilmesi gerekmektedir. Aynı işlem kelimelere ve hecelere ayırma sürecinde de yapılır. Bu nedenle yapısal analizi kelimelere ve hecelere ayırma süreci içerisinde yapmak metin işleme sürecinin performansını artıracaktır.

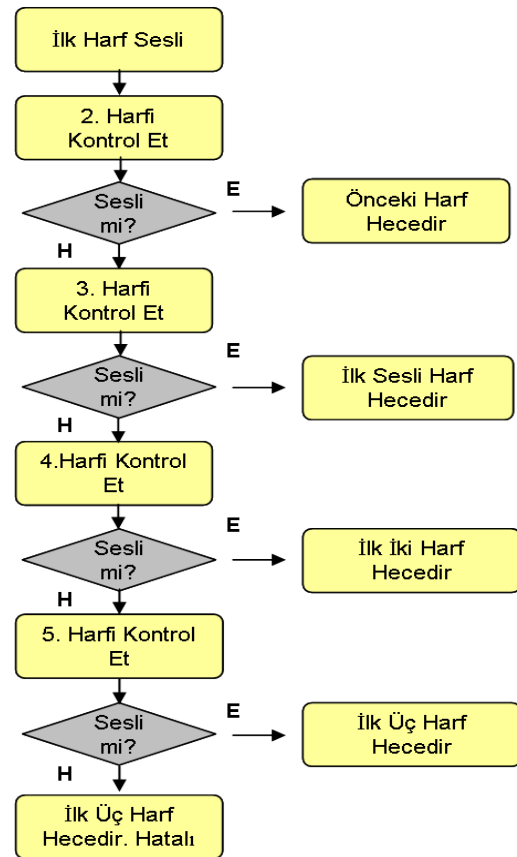
Vezinsel uyarılma sürecinde, kelimelerin ardışık olarak cümle içinde seslendirilirken uygun bir ezgi ölçüsünün olması amaçlanır. Ölçü, kelimeyi söylerken ağızdan çıkan seslerden ayrı olarak daha fazla konuşma özelliği içerir. Bunlar; zamanlama, duraksama, kelimeler üzerindeki vurgu gibi özelliklerdir. Doğru ölçü, doğru konuşma seslerini bulmak için ve doğru anlamı verebilmek için önemlidir. Örneğin; "Oku, baban gibi işsiz olma" cümlesi ile "Oku baban gibi, işsiz olma" cümlesi arasındaki anlam farkını yakalamak için duraksama ve zamanlama önemlidir. Noktalama işaretleri haricinde, bir cümleyi doğru olarak ifade edebilmek, doğru vurgulama yapabilmek için cümlelerin anlamını anlamak gereklidir ve maalesef bilgisayarlar bunu yapamazlar.

Son olarak konuşmanın üretimi sürecinde, metin işleme ve yapısal analiz süreciyle hecelere ayrılmış, noktalama işaretleri yardımıyla vezinsel uyarılması yapılmış girdi metni, ses veri tabanı ile eşleştirilir.

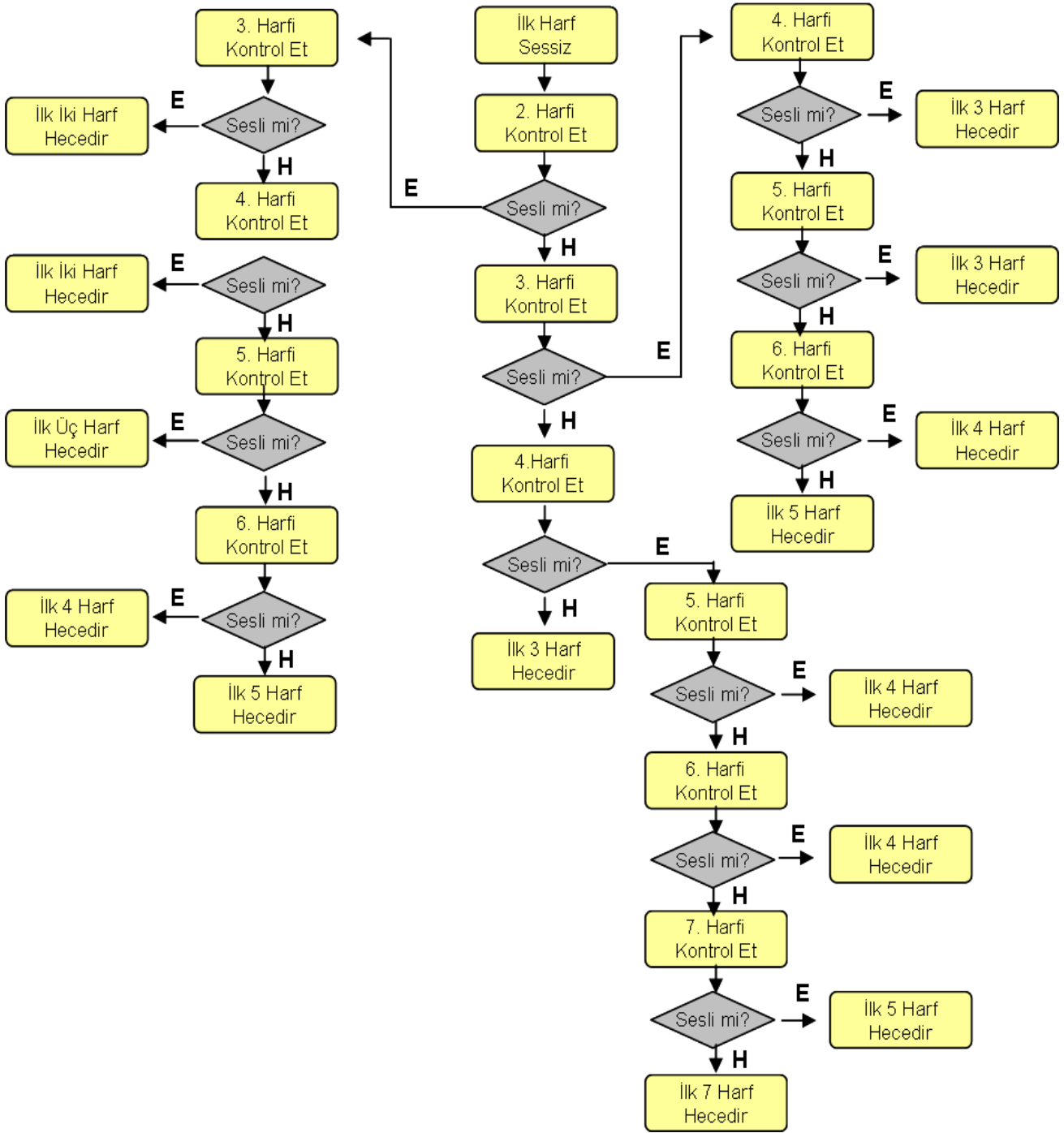
Eşleşen fonemlere ait ses birimleri ve duraksamalar birleştirilerek ses dalgası üretilir. Ses dalgası üretiminin birçok yolu vardır. En güncel sistemler genel olarak bunu iki yolla yaparlar. Bunlardan biri kayıtlı insan sesi parçalarını birleştirerek, diğeri ise sinyal işleme tekniklerini kullanarak yapar [9]. Bu yazılımda kaydedilen insan sesi ile bir ses veri tabanı oluşturulmuş ve bu ses veri tabanını kullanarak ses parçaları birleştirilmiştir.

3.2 Heceleme Algoritması

Türkçe kelimeleri hecelere ayırabilmek için her harfi bilmek gerekmez. Bu iş için, sesli harflerle sessiz harfleri birbirinden ayırt edebilmek yeterlidir. Örneğin her sesli yerine "a" harfi ve her bir sessiz yerine "b" harfi konursa, heceleme doğru olarak yapılabilir. Örneğin bu kurala göre "karanfil" sözcüğünün, sesli sessiz dağılımı "bababbab" olur. Bu heceleme algoritmasının ilk adımınıdır. Hecelere ayrılacak kelime sesli ve sessiz harf olarak ayrıştırıldıktan sonra, dilbilgisi kurallarına tabi tutulur. Mevcut kurallar doğrultusunda sesli harf ile başlayan bir kelimenin heceleme algoritması Şekil 6'da, sessiz harf ile başlayan bir kelimenin heceleme algoritması Şekil 7'de görülmektedir.



Şekil 6. Sesli harf ile başlayan bir kelimenin hece tespiti



Şekil 7. Sessiz harf ile başlayan bir kelimenin hece tespiti

Bu çalışma kapsamında hazırlanarak, geliştirilen yazılım içine entegre edilen Türkçe heceleme algoritması elde edilen sözcüklerin hecelere ayrılmasını sağlamaktadır. Örnek olarak “karanfil” sözcüğü incelenecek olursa; ilk olarak sözcüğün ilk harfinin sesli veya sessiz olup olmadığına bakılır. İlk harf “k” harfi olduğu için sessiz harftir. İkinci harfin sesli-sessiz durumuna bakılır. İkinci harf “a” olduğu için sesli harftir. Oluşan “k” ve “a” harflerinin hece olup olmadığını ardından gelen harfler belirleyecektir. Üçüncü harfe bakılır. Üçüncü harf sesli harf olsaydı “k” ve “a” harfleri “ka” hecesini oluşturacaktı. Örneğin “ka-os” gibi. Üçüncü harf olan “r” harfi sessiz harf olduğu için “k”, “a” ve “r” harflerinin hece olup olmadığını ardından gelen harf belirleyecektir.

Dördüncü harfe bakılır. Dördüncü harf sessiz harf olsaydı, heceyi belirleme işi beşinci harfe düşecekti. Beşinci harf sesli harf olsaydı örneğin “karga” “k”, “a” ve “r” harfleri “kar” hecesini oluşturacaktı. Hiç beşinci harf olmasaydı Örneğin “kars” gibi; ilk dört harf olduğu gibi hece olacaktır. Dördüncü harf olan “a” harfi sesli harf olduğu için “k”, “a” harfleri “ka” hecesini oluşturmaktadır. Oluşan “ka” hecesi kelimedenden kesilerek hece dizisine eklenir ve ilk hece elde edilmiş olur. Kelimedenden geriye kalan “ranfil” kelimesi öz yinelemeli olarak yeniden algoritmaya girer. İkinci kez Türkçe heceleme kurallarına göre değerlendirilerek elde edilen “ran” hecesi de kesilerek hece dizisine eklenir. Geriye kalan “fil” kelimesi

için aynı adımlar tekrar edilerek “fil” hecesi elde edilir ve “karanfil” kelimesi, “ka-ran-fil” olarak hecelenmiş olur.

Geliştirilen heceleme algoritmasının kelimeleri dil bilgisi kurallarına uygun olarak heceleylebildiğini ölçmek için bir doğrulama çalışması gerçekleştirilmiştir. Türkçe İmla Kılavuzunda yer alan kelime ve ifadeler programın imkân ve kabiliyeti doğrultusunda hecelere ayrılmıştır. Bu kapsamda rastsal olarak seçilen, program tarafından hecelere ayrılmış 10 000 adet kelimenin doğru olarak hecelere ayrılıp ayrılmadığı manüel olarak incelenmiştir. İmla Kılavuzunda yer alan kelimelerden örnek olarak alınan farklı hece yapısındaki bazı kelimelerin geliştirilen heceleme algoritması ile hecelenmiş halleri Tablo 4’te verilmiştir.

Tablo 4. Farklı kelimelerin geliştirilen yazılım tarafından hecelenmiş halleri

Kelime Yazılışı	Hecelenmiş Hali
alaturkacılık	a-la-tur-ka-cı-lık
alelacele	a-le-lâ-ce-le
bronzlaşmak	bronz-laş-mak
cümbür cemaat	cüm-bür ce-ma-at
çingiraklı yılançiller	çin-gi-rak-lı yı-lan-gil-ler
dezenfektan	de-zen-fek-tan
elektroansfalografi	e-lekt-ro-an-se-fa-log-ra-fi
elhamdülillah	el-ham-dü-lil-lâh
erkâniharbiye umumiye	er-kâ-ni-har-bi-ye-i u-mu-mi-ye
entelektüalizm	en-te-lek-tü-a-lizm
flândra balığı	fland-ra ba-lı-ğı
genel grev	ge-nel grev
Helenist	he-le-nist
ırk bilimi	ırk bi-li-mi
iadeli taahhütü	i-a-de-li ta-ah-hüt-ü

Geliştirilen yazılım tarafından heceleme işlemine tabi tutulan 10 000 kelimenin, ayrıca manüel olarak yapılan kontrolünde tamamının dil bilgisi kurallarına uygun bir şekilde hecelendiği tespit edilmiştir. Geliştirilen heceleme algoritması % 100 oranında doğru olarak çalışmaktadır.

4. BULGULAR VE YORUMLAR

Geliştirilen konuşma sentezleme yazılımında; sistemin temel amacı, girdi metinlerinin insan sesine yakın bir ses ile okutulmasıdır. Bu sebeple geliştirilen yazılımın ses kalitesinin ve sesin anlaşılır olup olmadığının test edilmesine ihtiyaç duyulmuştur. Bu testin amacı; geliştirilen yazılımın, geçerlemesi ve doğrulaması aşamasında gereksinimleri karşılayıp karşılamadığını anlamaktır. Bu maksatla geliştirilen yazılıma bir test modülü eklenmiştir. Bu test modülü yazılıp içine gömülü olarak 10 adet farklı tip sorudan oluşmaktadır. Sorular,

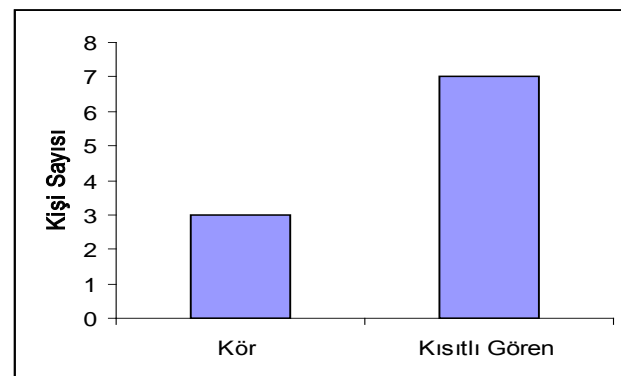
geliştirilen konuşma sentezleme yazılımı tarafından okunmuştur.

Hazırlanan 10 adet sorunun ilk altısı basit düzeyde genel kültür soruları olup, teste katılan görme engelli kişinin yazılım tarafından üretilen konuşmayı anlayıp anlamadıklarını ölçmeyi amaçlamaktadır. Kişinin soruya doğru veya yanlış cevap vermesi önemli değildir. Temel amaç, soruların anlaşılabilirliğini ölçmektir. İlk altı sorunun genel kültür sorusu olarak seçilmesinin sebebi, teste katılan kişinin konuşmayı daha dikkatli dinlemesini sağlamaktır. Diğer dört soru, ilk altı sorudaki amacı gütmekle birlikte teste katılan görme engelli kişinin bilgisayar kullanım seviyesini tespit etmeyi, hazırlanan yazılım ve yazılımın ürettiği ses ile ilgili düşüncelerini almayı amaçlamaktadır. Yapılan testte hazırlanan 10 adet soru Tablo 5’te verilmiştir.

Tablo 5. Test süresince sorulan sorular

S.Nu.	Soru
1	Türkiye’deki il sayısı nedir?
2	En büyük ilimiz hangisidir?
3	Atatürk hangi yılda doğmuştur?
4	Cumhuriyetin ilanı ne zaman olmuştur?
5	19 sayısının rakamlarını değiştirdiğimizde hangi sayıyı elde ederiz?
6	Matematikte toplamadaki etkisiz eleman hangi sayıdır?
7	Bilgisayar kullanım seviyeniz nedir?
8	Görme engelinize yönelik nasıl bir program yapılmasını isterdiniz?
9	Görme engellilere yönelik hazırlanacak bir kitap okuma programını faydalı buluyor musunuz?
10	Dinlediğiniz programdaki sesin kalitesini çok kötü, kötü, orta, iyi, çok iyi olarak ifade ediniz.

Test işlemleri Ankara Büyükşehir Belediyesine bağlı olarak Beşevler’de hizmet veren Görme Engelliler Eğitim ve Teknoloji Merkezinde yapılmıştır. Teste 10 görme engelli kişi katılmıştır. Bunların 8’i erkek 2’si bayandır. Bu kişilerin yaşı 10 ile 46 arasındadır. Şekil 8’de engel durumları gösterilmektedir.



Şekil 8. Görme kaybı durumları grafiği

Soruların tamamının katılımcılar tarafından anlaşılabilir ve kaçınıcı tekrardan sonra anlaşılabilirliğine ilişkin istatistik tutulmuştur. Hazırlanan istatistik, Tablo 6’da görülmektedir. Çizelge incelendiğinde, soruların

kullanıcılar tarafından ağırlıklı olarak tek seferde anlaşıldığı görülmektedir. Katılımcılar genellikle ilk sorularda sese alışma süreci yaşamışlar ve 3. ve sonraki sorularda soruları tek seferde anlamışlardır.

Tablo 6. Katılımcıların soruları dinleme sayıları

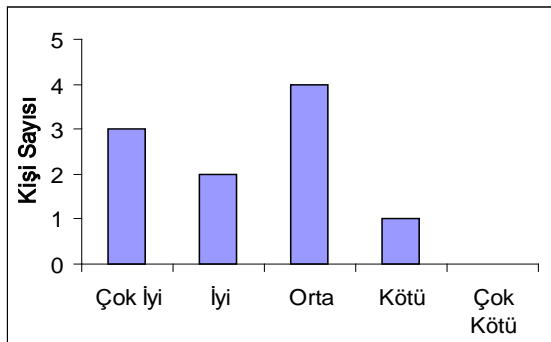
Katılımcıların Soruları Dinleme Sayıları										
Katılımcı \ Soru	1	2	3	4	5	6	7	8	9	10
1.Katılımcı	2	1	1	1	1	1	1	1	1	1
2.Katılımcı	1	1	1	1	1	1	1	1	1	1
3.Katılımcı	1	1	1	1	1	1	1	1	1	1
4.Katılımcı	2	2	1	1	2	1	1	1	1	1
5.Katılımcı	1	1	1	1	1	1	1	1	1	1
6.Katılımcı	2	1	1	1	1	1	1	1	1	1
7.Katılımcı	1	1	1	1	1	1	1	1	1	1
8.Katılımcı	2	2	1	1	1	1	1	1	1	1
9.Katılımcı	1	1	1	1	1	1	1	1	1	1
10.Katılımcı	1	1	1	1	1	1	1	1	1	1

9.soruda katılımcılara, görme engellilere yönelik hazırlanacak bir kitap okuma programını faydalı bulup bulmadıkları sorulmuştur. Katılımcıların tümü bu soruya faydalı bulduklarını ve yapılan çalışmanın özellikle ders çalışma, sınava hazırlık gibi süreçlerine katkı sağlayacaklarını belirtmişlerdir.

Son soruda katılımcılara dinledikleri programdaki sesin kalitesi sorulmuş ve çok kötü, kötü, orta, iyi, çok iyi olarak puan vermeleri istenmiştir. Programdaki konuşmanın kalite grafiği Şekil 9'da ve alınan cevaplara göre puanlama bilgileri Tablo 7'de verilmiştir.

Tablo 7. Programdaki konuşma kalitesinin puanlanması

Puan	Programdaki Sesin Kalitesi	Kişi Sayısı
5	Çok İyi	3
4	İyi	2
3	Orta	4
2	Kötü	1
1	Çok Kötü	0



Şekil 9. Programdaki konuşmanın kalite grafiği

Programda üretilen sesin kalitesi MOS (Mean Opinion Score) benzeri puanlama ile ölçülmeye çalışılmıştır. MOS subjektif ve ilk olarak PSTN (Public Switched Telephone Network) için sesin kalitesinin ölçülebilmesi için geliştirilmiş bir metottur. İstatistiksel anlamda önemli sayılabilecek bir grubun ses örneklerini 0 ile 5 arasında

puanlamasıyla oluşmaktadır ve böylece sesin istenen kalitede olup olmadığına karar verilir.

Yapılan hesaplamada çok iyiden çok kötüye toplam 5 seçenekte katılımcılardan konuşmanın kalitesine not vermeleri istenmiştir. Seçeneklerin puanlama şekli ve kişi sayısı MOS'a uygun olarak Tablo 7'de verilmiştir. Toplam 10 katılımcı tarafından verilen puanların toplamı 37 olmuştur. Verilen puanların toplamı katılımcı sayısına bölündüğünde puanların ortalaması 3,7 olmuştur.

5. SONUÇLAR

Konuşma tanıma ve konuşma sentezleme teknolojileri görme engelli kullanıcılara büyük kolaylıklar sağlamaktadır. Konuşma sentezleme teknolojileri sayesinde görme engelli kullanıcılar bilgisayar ekranında göremedikleri yazıları, konuşma sentezleyici yazılımlar sayesinde sesli olarak duyabilmektedirler. Ekran okuyucu programlar görme engellilerin bilgisayar kullanımlarında en büyük yardımcılarıdır.

Bu çalışma kapsamında, konuşma sentezleme teknolojileri konusunda detaylı araştırmalar yapılmıştır. Edinilen tecrübeler doğrultusunda Türkçe Hece Tabanlı Metinden Konuşma Sentezleme yapabilen bir elektronik kitap okuyucu yazılımı geliştirilmiştir. Teknik olarak parça tabanlı ardışık eklemeli(concatenative) konuşma sentezleme tekniği kullanılmıştır.

Geliştirilen MKS yazılımı içine entegre edilebilen bir "Türkçe Heceleme Algoritması" geliştirilmiştir. Algoritma, elde edilen kelimelerin hecelere ayrılmasını sağlamaktadır. Geliştirilen algoritmanın %100 oranında doğru olarak çalıştığı test edilmiştir.

Geliştirilen MKS yazılımında Türkçe dil yapısına uyumlu daha farklı ve daha pratik bir kesimleme ve etiketleme yöntemi sunulmuştur. Tüm hece birleşimlerinin ses kaydından kesimlenmesi yerine sadece bir ve iki harfli hece birleşimleri seslendirilerek hecenin başlangıç ve bitiş sınırları belirlenmiş ve ses veri tabanına hece ismiyle etiketlenmiştir. İki büyük harfli heceler yazılımın çalışma anında önce ikili hece formatına dönüştürülerek uygun ikili hece etiketi belirlenmekte, ardından fazladan harf kaldırılarak uygun ses üretilmektedir. Bu yöntemle herhangi bir sıkıştırma algoritması kullanmadan yaklaşık 5 GB'lık yer kaplayacak olan ses veri tabanı 10 MB'a düşürülmüştür. Bunun yanında, kesimleme ve etiketleme süreci de zaman açısından çok ciddi oranda azaltılmıştır.

Çeşitli görme kaybı olan 10 görme engelli tarafından geliştirilen yazılım test edilmiştir. Test sonucunda, yazılım tarafından üretilen konuşmanın kalitesi, kullanıcı değerlendirmesine göre, MOS benzeri bir derecelendirmeden 5 üzerinden 3,7 puan almıştır. MOS puanları son kullanıcıların düşüncesiyle yakından alakalıdır ve verilen puanların ortalaması 3,5 ve üzeriyse sesin kabul edilebilir düzeyde iyi olduğu anlaşılır.

Katılımcılar, geliştirilen yazılımda dinledikleri sesi, sürekli kullandıkları ve alışık oldukları ekran okuyucu

programındaki Gül sesi ile kıyasladıklarını belirtmişlerdir. Sesin biraz daha seri okunması halinde daha iyi olacağını ve bazı 3 harfli hecelerde ilk harfte vurgunun az yapıldığını ifade eden katılımcılar, ancak sesin mekanik olmadığını ve insan sesine yakın bir konuşma yapıldığını belirtmişlerdir.

Geliştirilen yazılımda elde edilen konuşma çıktısının anlaşılır olduğu değerlendirilmiştir. Ancak daha kaliteli bir konuşma elde edilebilmesi için ses veri tabanının oluşturulması aşamasında, ses birimlerinin kesimlenmesi daha kaliteli bir ses kayıt ortamında yapılmalı ve ikili heceler; diksiyonu, ses tonlaması düzgün ve orantılı bir konuşmacı tarafından seslendirilmelidir. Ayrıca, daha vezinsel bir konuşma elde edilebilmesi için, ses veri tabanından ayrı olarak, anlamlı uzun bir metin okunup kayıt altına alınarak metindeki duraksama noktaları ve süreleri tespit edilmelidir. Elde edilen veriler bir vezinsel algoritmaya dönüştürülerek daha akıcı bir konuşma elde edilebilir.

KAYNAKLAR

- [1] R. Aşlıyan., K. Günel, "Türkçe metinler için hece tabanlı konuşma sentezleme sistemi", **Akademik Bilişim 2008**, Çanakkale Onsekiz Mart Üniversitesi, 31-38, 2007.
- [2] J. Holmes, W. Holmes, **Speech Synthesis and Recognition, 2nd Edition**, CRC Press, London, 2001.
- [3] D. Aksan, **Türkiye Türkçesinin Dünü, Bugünü, Yarını**, Bilgi, Ankara, 68, 2000.
- [4] N. Büyükkantarıcıoğlu, "Türkçe sözcük biçimlenmesinde düzlemler ve türetmeler", Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi, 17(1), 81-94, 2000.
- [5] İnternet: Türk Dil Kurumu, Büyük Türkçe Sözlük, <http://tdkterim.gov.tr/bts/>, 10.06.2011.
- [6] İnternet: Türk Dil Kurumu, Hece Yapısı ve Satır Sonunda Kelimelerin Bölünmesi, <http://www.tdk.gov.tr>, 01.06.2011.
- [7] E. Mengüşoğlu, **Bir türkçe sesli ifade tanıma sisteminin kural tabanlı tasarımı ve gerçekleştirimi**, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, 1999.
- [8] Milli Eğitim Bakanlığı Meslekî Eğitim ve Öğretim Sisteminin Güçlendirilmesi Projesi, **Mesleki Gelişim Diksiyon 1**, MEGEP Yayınları, Ankara, 2006.
- [9] İnternet: Blog Sayfası, Speech Synthesis Algorithms (Konuşma Sentezleme Algoritmaları), <http://gungorbasa.blogspot.com/2011/02/speech-synthesis-algorithmskonusma.html>, 26.05.2011.
- [10] R. Aşlıyan, K. Günel, A. Filiz, "Türkçe Otomatik Heceleme Sistemi ve Hece İstatistikleri", **Akademik Bilişim 2006, BilgiTek IV**, Pamukkale Üniversitesi, Denizli, 9-11 Şubat 2006.
- [11] A.E. Yılmaz, "Türkçe metinden konuşma sentezleme uygulamaları için bir veri sözlük seti ve yazılım çerçevesi", Gazi Üniv. Müh. Mim. Fak. Dergisi, 24(4): 735-744, 2009.
- [12] J. Zhang, **Language generation and speech synthesis in dialogues for language learning**, Y Yüksek Lisans Tezi, Department of Electrical Engineering and Computer Science, Massachusetts, 2004.
- [13] J. Kominek, A.W. Black, "CMU ARCTIC databases for speech synthesis", Language Technologies Institute, 3-7, 2003.
- [14] İ. Y. Özüm, **A speech synthesis system for turkish language based on the concetanation of phonemes taken from speaker**, Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Fen Bilimleri Enstitüsü, 1993.
- [15] K. Pohlmann, **Principles of Digital Audio, 3rd Edition**, ISBN 0-07-050468-7, McGraw-Hill, Inc., New York, 1995.
- [16] P.H. Santen, R.W. Sproat, J.P. Olive, J. Hirschberg, **Progress in Speech Synthesis**, Springer Verlag, New York, 1996.
- [17] H. Sak, **A corpus based concenative speech synthesis system for turkish**, Yüksek Lisans Tezi, Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü, 2004.
- [18] J. Allen, S. Hunnicutt, D. Klatt, **From Text to Speech: The MITalk System**, Cambridge University Press, New York, 1987.