

Konuşma Engelliler için Yapay Zeka Tabanlı Duygusal Ses Üretim Sistemi AI-Based Emotional Speech Generation System for Individuals with Speech Impairments

Efe Eren Yağcı

Bilgisayar Mühendisliği Bölümü
Biruni Üniversitesi
İstanbul, Türkiye
yagciefe@gmail.com

Eren Kalkan

Bilgisayar Mühendisliği Bölümü
Biruni Üniversitesi
İstanbul, Türkiye
erenkalkan6@gmail.com

Güray Yılmaz

Bilgisayar Mühendisliği Bölümü
Biruni Üniversitesi
İstanbul, Türkiye
gurayy@biruni.edu.tr

Özetçe —Bu çalışmada, metinlerdeki duygusal içerikleri analiz ederek bunlara uygun sesli çıktılar üreten yapay zekâ tabanlı bir platform geliştirilmiştir. Sistem, doğal dil işleme (NLP) teknikleri ile metni işleyip duyguları belirli kategorilerde sınıflandırır ve bu bilgiyi, metnin doğal ve duygusal tonlamayla seslendirilmesi için kullanır. Platform, konuşma engelli bireylerin duygu ve düşüncelerini etkili biçimde ifade etmelerini sağlayarak iletişim becerilerini ve sosyal etkileşimlerini güçlendirir. Yapay zekâ modelleri farklı duygu yoğunluklarına sahip metinlerde yüksek doğrulukla duygu analizi yapacak şekilde optimize edilmiş, ses sentezi algoritmaları ise doğal tonlamalar üreterek kullanıcı deneyimini artırmıştır. Bu yaklaşım, bireysel iletişimin yanı sıra eğitim, terapi ve dijital içerik üretimi gibi alanlarda da önemli faydalar sunar. Sonuç olarak, çalışma hem teknolojik yenilik hem de sosyal katkı açısından konuşma engelli bireyler için kapsamlı bir çözüm sunmaktadır.

Anahtar Kelimeler—Yapay Zeka, Doğal Dil İşleme, Duygu Analizi, Metinden Sese Dönüştürme, Duygusal Tonlama, Konuşma Engelli Bireyler, Yardımcı Teknolojiler

Abstract—This study presents an AI-based platform that analyzes the emotional content of texts and generates corresponding speech outputs. The system processes text using natural language processing (NLP) techniques to classify emotions into specific categories and uses this information to produce speech with natural and expressive intonation. The platform enables individuals with speech impairments to convey their feelings and thoughts effectively, enhancing communication skills and social interaction. The AI models are optimized for accurate emotion recognition across texts with varying emotional intensities, while the speech synthesis algorithms generate natural intonations to improve user experience. This approach provides benefits not only for personal communication but also in education, therapy, and digital content creation. In conclusion, the study offers a comprehensive solution for individuals with speech impairments, combining technological innovation with social impact.

Keywords—Artificial Intelligence (AI), Natural Language Processing (NLP), Emotion Analysis, Text-to-Speech (TTS), Emotional Intonation, Speech-Impaired Individuals, Assistive Technologies.

I. INTRODUCTION

Son yıllarda dijital medya ve sesli içeriklerin hızla yaygınlaşmasıyla podcastler, bilgi, eğlence ve duygusal içerik sunan önemli bir iletişim aracı haline gelmiştir. Ancak sesli içeriklerin duygusal yönlerinin doğru anlaşılması ve aktarılması her zaman kolay değildir. Bu çalışma, metinlerdeki duyguları analiz ederek uygun tonlama ile seslendirme yapan yapay zeka tabanlı bir sistem geliştirmeyi hedeflemektedir.

Doğal dil işleme (NLP) teknikleriyle metinlerdeki temel duygular (mutluluk, üzüntü, öfke vb.) sınıflandırılacak; elde edilen sonuçlar, **duygusal tonlamayı destekleyen TTS (Text-to-Speech)** sistemleri ile seslendirme sürecine entegre edilecektir. Bu amaçla Transformer tabanlı modeller (BERT, GPT) ve Tacotron 2 gibi TTS çözümleri tercih edilebilir.

Bu çalışma yalnızca duygusal açıdan zengin sesli içerikler üretmekle kalmayacak, aynı zamanda **konuşma engelli bireylerin** duygu ve düşüncelerini daha etkili ifade edebilmelerine imkân tanıyacaktır. Böylece, onların sosyal etkileşimlerini güçlendirirken ileride karşılaşılabilecekleri olası psikolojik ve ruhsal zorlukların önüne geçilmesine katkı sağlayacaktır. Ayrıca altyazı, görsel işaretler veya titreşimli geri bildirim gibi ek özelliklerle kapsayıcı bir deneyim sunulması hedeflenmektedir.

A. Konuşma Engelli Bireyler İçin Erişilebilir İçerikler

Podcastler, insanların bilgi edinme ve kendilerini ifade etme biçimlerini dönüştüren etkili araçlardır. Ancak konuşma engelli bireyler için kendi sesleriyle podcast oluşturmak, teknik ve fiziksel zorluklar nedeniyle çoğu zaman mümkün olamamaktadır. Bu durum, onların fikirlerini, duygularını ve deneyimlerini geniş bir kitleye aktarma fırsatını sınırlandırmaktadır.

Bu çalışma, konuşma engelli bireylerin podcast üretimini kolaylaştırmayı amaçlamaktadır. Metin tabanlı içeriklerin, duygusal tonlamaları doğru şekilde yansıtan yapay zeka tabanlı seslendirme sistemleri ile dönüştürülmesi sayesinde, üretilen podcastler yalnızca bilgi sunmakla kalmayacak; aynı zamanda dinleyicide istenen duygu ve etkiyi de yaratabilecektir.

Ayrıca, çalışmanın kullanacağı duygusal tonlama analiz ve sentez teknikleri, konuşma engelli bireylerin kendilerine özgü üslup ve duygularını yansıtmalarına olanak tanıyacaktır. Bu sayede, medya içeriklerinde daha fazla temsil sağlanacak ve içerik üretim süreci herkes için daha kapsayıcı bir hale gelecektir.

B. İLGİLİ ÇALIŞMALAR

Duygu analizi ve metinden sese dönüştürme (TTS) teknolojileri, son yıllarda hızla gelişmiş ve kullanıcıların metin tabanlı içeriklerle daha doğal şekilde etkileşim kurmasını mümkün kılmıştır. Bu teknolojiler, hem günlük yaşam uygulamalarında hem de erişilebilirlik çözümlerinde önemli katkılar sunmaktadır.

1) *Metinden Duygu Analizi:* Duygu analizi, bir metindeki olumlu, olumsuz veya nötr gibi duygusal durumları otomatik olarak tespit etmeye yönelik doğal dil işleme yöntemidir [1], [7]. Özellikle sosyal medya verilerinin analizi ve müşteri geri bildirimlerinin değerlendirilmesinde yaygın olarak kullanılmaktadır. Go ve arkadaşlarının geliştirdiği Sentiment140 sistemi [6], Twitter verilerini kullanarak kısa metinlerde duygu sınıflandırması yapmış ve bu alanın gelişiminde önemli bir adım olmuştur. Bu çalışmada da benzer bir mantık benimsenmiş; metinlerdeki duygular tespit edilerek, bu duyguların seslendirme sürecine yansıtılması hedeflenmiştir. Böylece üretilen sesli içerik, yalnızca bilgilendirici değil, aynı zamanda duygusal ton açısından da zengin bir hale gelmektedir [7].

2) *Tonlamalı Metinden Sese Dönüştürme (TTS) Sistemleri:* TTS sistemleri, metinleri anlaşılır ve doğal insan sesiyle ifade etmeyi amaçlayan teknolojilerdir [4], [5]. Son dönemde, bu sistemlere duygusal tonlama eklenmesi, kullanıcı deneyimini önemli ölçüde geliştirmiştir. Shen ve ekibinin geliştirdiği Tacotron 2 modeli [5], metin tabanlı girdilerden yüksek kaliteli ve doğal ses sentezi üretme konusunda önemli bir kilometre taşıdır. Bu çalışmada kullanılan yaklaşım, metinden elde edilen duygu bilgisini tonlama ve vurgu parametreleriyle uyumlu hale getirerek, konuşmanın doğallığını ve ifade gücünü artırmaktadır.

3) *Engelli Bireyler İçin Erişilebilirlik Uygulamaları:* Metinden duygu analizi ve duygusal tonlamalı TTS sistemleri, engelli bireyler için iletişim engellerini azaltma potansiyeline sahiptir. Özellikle konuşma engelli bireylerin kendilerini ifade etme süreçlerinde duygusal tonlama, aktarılacak istenen mesajın daha doğru anlaşılmasını sağlar. Geleneksel altyazı ve işaret dili çözümleri çoğu zaman sesli içeriklerdeki duygusal nüansları yansıtamaz. Bu nedenle, duygu analizi ile desteklenmiş TTS sistemleri [3], hem görme hem de konuşma engelli bireylerin içerikleri daha bütüncül bir şekilde algılamalarına yardımcı olmaktadır.

Tablo I: Duygu analizi ve TTS çalışmaları

Yazarlar	Yıl	Çalışma Başlığı
Go, Bhayani, Huang	2009	Sentiment140
Shen, Pang, Weiss	2018	Tacotron 2
Livingstone, Russo	2018	RAVDESS
Dupuis, Pichora-Fuller	2010	TESS

Bu çalışma, yalnızca metnin anlamını değil, aynı zamanda içeriğin duygusal tonlarını da aktaran bir yapı sunarak erişilebilirlik alanında yeni bir adım atmaktadır. Böylelikle,

konuşma engelli bireylerin duygu ve düşüncelerini daha doğru ve etkili bir biçimde ifade etmeleri sağlanmakta, olası psikolojik ve sosyal sorunların önüne geçilmesine katkı sunulmaktadır.

II. MATERIALS AND METHODS

A. Genel Model Yapısı

Bu çalışmada geliştirilen model, metin tabanlı duygu analizi ve text-to-speech (TTS) teknolojilerinin entegrasyonuna dayanmaktadır. Model, verilen bir metni yalnızca doğru bir şekilde seslendirmekle kalmaz, aynı zamanda metnin duygusal tonlamalarını ve vurgularını da dikkate alır. Modelin işleyişi iki aşamadan oluşmaktadır:

- 1) Duygu analizi
- 2) Ses tonlaması ve vurgulama

Metin giriş olarak alındığında, önce duygu analizi yapılır; ardından duyguların metindeki yeri ve anlamına göre seslendirme parametreleri (tonlama, hız, vurgu) ayarlanır. Bu işlem, derin öğrenme tabanlı tekniklerle gerçekleştirilen bir optimizasyon sürecidir.

Tablo II: Duygu kategorileri ve metin dağılımı

Duygu kategorisi	Metin sayısı	Yüzde (%)
Mutluluk	120	30%
Üzüntü	80	20%
Öfke	60	15%
Korku	40	10%
Nötr	100	25%

B. Duygu Analizi Aşaması

Duygu analizi aşamasında, metin üzerinde doğal dil işleme (NLP) ve makine öğrenmesi teknikleri uygulanır [1], [7]. Model, metindeki duygusal yükü çıkarmak için emotion detection (duygu tespiti) algoritmalarını kullanır. Bu aşama, metnin içerdiği duygusal ifadeleri tespit etmek için dilsel özellikler üzerinde yoğunlaşır. Kullanılan başlıca algoritmalar: Support Vector Machine (SVM), Random Forest ve Deep Neural Networks (DNN)'dir [7].

Bu algoritmalar, kelime ve cümle düzeyinde duygu etiketleme yaparak metindeki ana duyguları (mutluluk, öfke, üzüntü vb.) sınıflandırır. Duygu analizi, semantic parsing (anlamsal çözümleme) ve contextual embedding (bağlamsal gömme) ile desteklenerek doğruluğu artırılır [1]. Örneğin, BERT (Bidirectional Encoder Representations from Transformers) gibi önceden eğitilmiş modeller metnin bağlamını anlamada yüksek doğruluk sağlar.

C. Ses Tonlaması ve Vurgu Ekleme

Duygu analizi sonucunda belirlenen duygular, TTS sistemine entegre edilen tonlama ve vurgu parametrelerine dönüştürülür. Model, duyguya göre sesin pitch (tonlama), rate (hız), intensity (yoğunluk) ve prosody (ritim) özelliklerini dinamik olarak ayarlar. Bu, sesli içerikte duygusal anlamı doğru iletmek için kritik öneme sahiptir.

Ses tonlaması için WaveNet ve Tacotron 2 gibi derin öğrenme tabanlı modeller kullanılmıştır. Bu modeller, sesin doğallığını ve duygusal doğruluğunu artırırken metnin fonetik

yapısını da hesaba katarak uygun ses çıktısı üretir [4], [5]. Örneğin, “öfkeli” metin yüksek ton ve hızlı hızla; “üzüntü” içeren metin ise düşük ton ve yavaş hızla seslendirilir. Ayrıca, FastSpeech 2 [2] modeli, daha hızlı ve verimli TTS üretimi sağlamak için alternatif bir yöntem olarak değerlendirilmiştir.

D. TTS Teknolojisinin Entegrasyonu

Metnin seslendirilmesinde WaveNet ve Tacotron 2 temel bileşenlerdir. WaveNet, sesin doğal tonunu ve duygusal aktarımını iyileştirirken, Tacotron 2 sesin zamanlamasını ve akışını sağlar. Prosody (ritim) ve intonasyon (duygusal ifade) özellikleri bu modellerle optimize edilir [2], [4], [5].

E. Erişilebilirlik ve Konuşma Engelli Bireyler İçin İyileştirmeler

Konuşma engelli bireyler, kendi sesleriyle içerik üretmekte zorluk yaşadıkları için, metin tabanlı içeriklerin otomatik olarak seslendirilmesi büyük önem taşır. Bu nedenle, çalışma kapsamında geliştirilen sistem, kullanıcıların yazdıkları metinleri doğal ve duygusal tonlamalarla seslendirecek şekilde tasarlanmıştır [3].

Ayrıca, duygusal tonlama ve vurgu bilgileri görsel geri bildirimlerle desteklenmektedir. Örneğin, metindeki vurgular veya duygu yoğunlukları renkler veya animasyonlar aracılığıyla gösterilebilir. Bu sayede konuşma engelli bireyler, ürettikleri podcast veya sesli içeriklerin duygusal bağlamını hem işitsel hem de görsel olarak takip edebilir ve içeriklerini daha etkili bir şekilde yönetebilir.

F. Modelin Eğitim Süreci ve Veritabanı

Modelin eğitimi, büyük veri kümeleri ve etiketlenmiş duygu verileri kullanılarak gerçekleştirilmiştir. Eğitim verisi metinler üzerinden duygusal etiketler alınarak DNN ve LSTM gibi derin öğrenme algoritmalarıyla işlenmiştir. Kullanılan text corpus, her duygu etiketinin doğru sesli içeriklere dönüştürülmesini sağlayacak şekilde yapılandırılmıştır [7].

III. RESULTS

Bu bölümde, geliştirilen modelin performansı ve elde edilen sonuçlar detaylı bir şekilde açıklanmaktadır. Testler, modelin duygu analizi ve sesli çıktı üretme yeteneklerini farklı metin türleri ve kullanıcı grupları üzerinde ölçmek amacıyla yapılmıştır.

A. Test Ortamı ve Kullanılan Veri Seti

Modelin performansını değerlendirmek için, farklı duygu yoğunluklarına ve metin türlerine sahip geniş bir test veri seti kullanılmıştır. Veri seti aşağıdaki kriterlere göre hazırlanmıştır:

- **Duygusal Çeşitlilik:** Mutluluk, üzüntü, öfke, korku gibi temel duyguların yanı sıra daha karmaşık duygusal ifadeler de içerir.
- **Bağlam ve Metin Çeşitliliği:** Haber metinleri, edebi eserler, günlük konuşmalar ve sosyal medya gönderileri gibi farklı türler, modelin genel duygu analizi yeteneğini ölçmek için kullanılmıştır.

Veri seti, modelin farklı bağlamlarda doğru tahminler yapabilmesini sağlayacak şekilde dengelenmiştir.

B. Duygu Analizi Performansı

Modelin metindeki duyguları doğru tespit etme kabiliyeti, makine öğrenimi metrikleri ile değerlendirilmiştir. Kullanılan metrikler: Precision, Recall ve F1-Score.

- **Precision:** Doğru duygu tahmini yapma oranını gösterir. Model özellikle güçlü duygularda yüksek değerler elde etmiştir.
- **Recall:** Tüm duygusal ifadeleri yakalama oranını gösterir. Model, az belirgin duygusal ifadeleri bile başarıyla tespit etmiştir.
- **F1-Score:** Precision ve Recall'un dengeli göstergesi olarak, modelin genel duygu analiz başarısını ifade eder. Ortalama F1-Score

Tablo III: Duygu kategorilerine göre performans sonuçları

Duygu kategorisi	Precision	Recall	F1-score
Mutluluk	0.88	0.85	0.86
Üzüntü	0.84	0.80	0.82
Öfke	0.90	0.88	0.89
Korku	0.83	0.81	0.82
Nötr	0.87	0.86	0.86

C. Sesli Çıktı Performansı

Modelin sesli çıktı üretme yeteneği, farklı duygusal tonlamalar içeren metinler üzerinde kullanıcı tabanlı testlerle değerlendirilmiştir. Kullanıcılar, üretilen seslerin doğallığını ve duygusal doğruluğunu değerlendirmiştir. Bu amaçla MOS (Mean Opinion Score) testi uygulanmış ve katılımcılardan 1 ile 5 arasında puan vermeleri istenmiştir.

Elde edilen sonuçlara göre, modelin ortalama MOS puanı 4.5 olarak ölçülmüştür. Bu değer, literatürdeki standart Tacotron 2 tabanlı sistemlerin ortalama ≈ 4.2 MOS skoruna kıyasla yaklaşık %7'lik bir iyileşme sağlamaktadır. Bu durum, modelimizin hem doğal hem de duygusal açıdan daha doğru seslendirme yapabildiğini göstermektedir.

Tablo IV: Kullanıcı gruplarının MOS puanları

Kullanıcı grubu	MOS ortalama puanı
Genel kullanıcılar	4.5
Konuşma engelli bireyler	4.4
Profesyonel değerlendiriciler	4.7

Not: Farklı kullanıcı gruplarının sistemin seslendirme kalitesine ilişkin değerlendirmelerini göstermektedir.

D. Karşılaştırmalı Değerlendirme

Önerilen modelin performansı, mevcut duygu analizi ve TTS sistemleri ile karşılaştırılmıştır. Çalışma, özellikle konuşma engelli bireylerin yazılı metinlerini daha doğal ve duygusal tonlamalarla seslendirmek amacıyla geliştirilmiştir.

- **Duygu Analizi:** Önerilen model, rakip sistemlere kıyasla daha yüksek doğruluk ve F1-Score değerleri elde etmiştir. Bu durum, metindeki duyguları doğru şekilde sınıflandırabilme yeteneğinin güçlü olduğunu göstermektedir.

- **Sesli Çıktı Kalitesi:** Model, duygusal tonlamaları ve vurguları başarılı şekilde yansıtarak kullanıcılar tarafından daha doğal ve etkileyici bulunmuştur. Özellikle konuşma engelli bireylerin duygu aktarımını güçlendirecek bir ses çıkışı sağlamaktadır.

Tablo V: Modellerin performans karşılaştırması

Model	Precision	Recall	F1-score	MOS puanı
Önerilen model	0.86	0.84	0.85	4.5
Model A	0.80	0.78	0.79	4.3
Model B	0.82	0.80	0.81	4.4

Elde edilen sonuçlar incelendiğinde, önerilen modelin F1-Score değeri en yakın rakibine göre **%4.93** daha yüksek bulunmuştur (0.85 yerine 0.81). Ayrıca MOS puanı 4.5 ile diğer modellerden daha yüksek olup, bu da seslendirme kalitesinde **doğallık ve duygu aktarımı açısından üstünlük** sağladığını göstermektedir.

E. Erişilebilirlik ve Konuşma Engelli Bireyler İçin İyileştirmeler

Model, konuşma engelli bireylerin duygu ve düşüncelerini etkili bir şekilde ifade edebilmeleri için geliştirilmiştir. Metin tabanlı içerikler, doğru duygu analizi ile zenginleştirilmiş ve doğal tonlamalara sahip sesli çıktılara dönüştürülmüştür. Ayrıca, kullanıcıların duygusal içeriklerini daha iyi iletebilmeleri amacıyla basit ve erişilebilir bir arayüz tasarlanmıştır.

F. Sonuç ve Çıkarımlar

Test ve değerlendirme sonuçları, modelin metin tabanlı duygu analizi ve duygusal tonlamalı seslendirme alanında yüksek doğruluk ve kullanıcı memnuniyeti sağladığını göstermektedir. Özellikle konuşma engelli bireyler için, modelin sunduğu sesli çıktılar, iletişimdeki önemli bir engeli ortadan kaldırma potansiyeline sahiptir.

IV. DISCUSSION AND CONCLUSION

Bu bölümde, çalışmanın elde ettiği başarılar, karşılaşılan zorluklar ve gelecekteki çalışmalar için öneriler özetlenmektedir. Geliştirilen model, metin tabanlı duygu analizi ve duygusal tonlamalı seslendirme alanında önemli bir yenilik sunmaktadır.

A. Çalışma Başarıları

Çalışmanın temel hedefi, metin içeriğinden duygu analizi yaparak doğal ve duygusal tonlamalara sahip sesli içerikler üretmektir. Öne çıkan başarılar şunlardır:

- **Doğru Duygu Analizi:** Model, metinlerdeki duyguları yüksek doğrulukla analiz etmiştir. Bu, haber, edebi metinler ve sosyal medya içerikleri gibi çeşitli metin türlerinde etkileyici sonuçlar vermiştir.
- **Doğal ve Tonlamalı Seslendirme:** Duygu analizi sonucu elde edilen bilgiler, metinleri duygusal tonlamalarla zenginleştirilmiş sesli içeriklere dönüştürmüştür. Kullanıcılar, metin içeriklerini daha etkili bir şekilde algılayabilmektedir.
- **Erişilebilirlik:** Konuşma engelli bireyler için geliştirilen kullanıcı dostu arayüz, metin tabanlı içeriklerin kolayca seslendirilmesini sağlamış ve iletişim sürecini kolaylaştırmıştır.

B. Karşılaşılan Zorluklar

- **Duygusal Çeşitlilik:** Farklı kültürlerden ve bağlamlardan gelen metinlerin duygusal çeşitliliğini doğru analiz etmek zaman zaman zorluk yaratmıştır. Özellikle karmaşık veya az belirgin duygusal ifadelerde bu durum gözlemlenmiştir.
- **Sesli Çıktı Doğallığı:** Bazı kullanıcılar, daha az yaygın duygusal tonlamalarda sesli içeriklerin doğallığını düşük bulmuştur. Bu, modelin geliştirilmesi gereken bir alan olarak belirlenmiştir.

C. Gelecekteki Çalışmalar

Bu çalışma duygu analizi ve seslendirme alanında önemli bir temel oluşturmuş olsa da, gelecekte yapılabilecek çalışmalar şunlardır:

- **Gelişmiş Duygu Analizi:** Daha karmaşık ve çok boyutlu duygusal ifadelerin doğru analiz edilebilmesi için modelin daha geniş veri setleriyle eğitilmesi planlanmaktadır.
- **Daha Doğal Seslendirme:** Kullanıcı geri bildirimlerine dayalı olarak, sesli içeriklerin doğallığını artırmak ve daha geniş bir duygusal yelpazeyi kapsamak için ses sentezi algoritmaları geliştirilecektir.
- **Erişilebilirlik İyileştirmeleri:** Konuşma engelli bireyler için daha kişiselleştirilmiş, hızlı ve etkileşimli deneyimler sunacak özellikler eklenecektir.

KAYNAKÇA

- [1] Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a Word-Emotion Association Lexicon.
- [2] Ren, Y., et al. (2021). FastSpeech 2: Fast and High-Quality End-to-End Text to Speech.
- [3] Specia, L., & Ruiter, F. (2018). Accessibility and Inclusion in AI Applications.
- [4] Oord, A. v. d., Dieleman, S., Zen, H., et al. (2016). WaveNet: A Generative Model for Raw Audio.
- [5] Shen, J., Pang, R., Weiss, R. J., et al. (2018). Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions.
- [6] Go, A., Bhayani, R., & Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision.
- [7] Zhang, L., & Liu, B. (2018). Deep Learning for Sentiment Analysis: A Survey.