

# ELM368 – DÖNEM PROJESİ

## Ses İşleme Yöntemiyle Cinsiyet Ayrımının Gerçekleştirilmesi

Arda Tan GÖKER, Gizem BAKAN, Mustafa Oğuz ÖZCAN  
1801022042, 1801022047, 1801022032  
a.goker2018@gtu.edu.tr, g.bakan2018@gtu.edu.tr, m.ozcan2018@gtu.edu.tr

### ABSTRACT (ÖZET)

Ses analizi, makine insan etkileşiminde önemli bir role sahiptir. Çevre, aksan ve diksiyon gibi farklı etkenler farklı şekilde kurgulanmış ses tanımlama sistemlerinin etkinliği etkileyen ortak faktörlerden bazılarıdır. Buna göre uygun filtreleme aşamaları gerçekleştirilmeli, söz konusu sesin asıl olarak ihtiva ettiği bilgi gürültülerden arındırılmalıdır. Ses analizinin özelinde kadın ve erkek cinsiyetlerine ait sesin işlenmesi, her iki cinsiyete ait seslerde karakter çıkarımının gerçekleştirilmesini içerir.

İnsan sesinde formant olarak adlandırılan belirli frekanslar etrafında yoğunlaşmış enerji bölgeleri bulunur. Kadın ve erkek ses tellerinin fizyolojik farklılıklarından kaynaklı olarak formant analizi, kadın ve erkek ses sinyallerinin karakterize edilip sınıflandırılabilmesine olanak sağlar. Bu nedenle bu makalede ses analizi, her iki cinsiyete ait en az onar saniyelik üç ses kaydının Fourier dönüşümü yöntemiyle frekans domainine çekilmesi içermektedir. Formant analizi hakkında bildiklerimizle örtüşecek şekilde, kadın ve erkek bireylere ait Fourier analizlerinde yoğunluklu görülen frekans bileşenlerinin farklı olması beklenmektedir. Buna göre sesi özellikle karakterize eden frekanslar filtrelenerek kadın ve erkek seslerinin ayrımının gerçekleştirilebilmesine çalışılacaktır.

### ANAHTAR KELİMELELER

FIR Filtre, IIR Filtre, Formant, Fourier, Pitch, Ses

## 1. Giriş

Bir konuşma sinyali, konuşmayı gerçekleştiren kişi hakkında bir dizi gerekli bilgiyi içerir. Akustik olarak uyarıcı ağız ve burun boşlukları tarafından üretilen ses/konuşma sinyali, kişiye ait özellikleri tanımlamak için kullanılabilir. Cinsiyet ayrımının gerçekleştirilmesi özelinde bu teknolojinin faydaları, herhangi bir bireysel ekipmana ihtiyaç duymadığı için pahalı olmamasıdır. Böyle bir konuşma sinyalinin yakalanabilmesi için bir sensör veya - bir veri seti üzerinden analiz yapılmak isteniyorsa - bir veri seti yeterlidir, ayrıca kullanılan

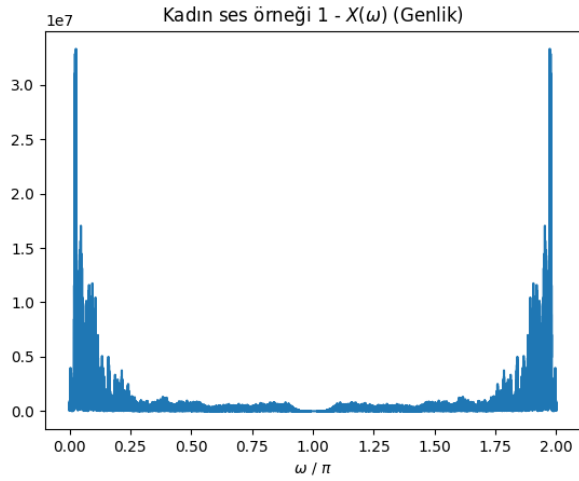
algoritmalar Fast Fourier Transform (FFT) gibi temel tekniklerin kullanılması ile kolayca kurgulanabilir.[1] Fast Fourier Transform, dijital sinyal işlemede sinyallerin spektrumlarının çıkarılması için kullanılan bir tekniktir ve karmaşık dönüşümlerin hesaplama süresini azaltan bir Fourier Transform gerçekleştirme algoritmasıdır. [2]

## 2. Deneyler ve Analiz

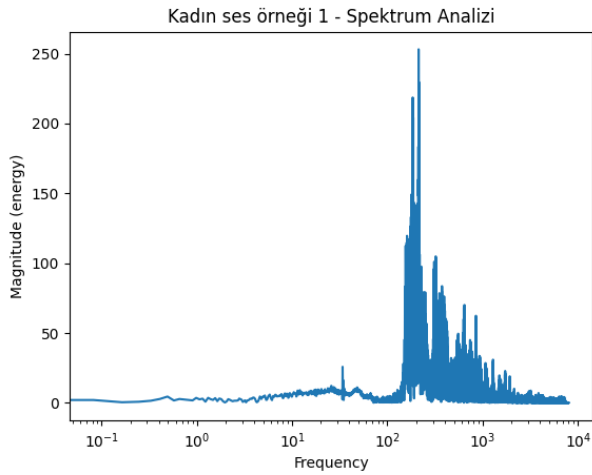
Öncelikle, kadın ve erkek sesinin ayrımının yapılabilmesi için aralarındaki farklılıkların tespit edilmesi, bu seslerin temel frekans içeriklerinin öğrenilmesi gerekmektedir. Bu işlem incelenen seslerin frekans uzayına dönüştürülmesini sağlayan Fourier dönüşümü ile yapılır. Frekans uzayındaki belirli bileşenler sesin karakteristiği hakkında çeşitli bilgiler verir.

### 2.1. Frekans Analizi

Üzerinde çalışılan ses kayıtları, internet üzerindeki bir bilimsel veritabanından alınmış olup, ikisi erkek ikisi kadın 4 farklı kişi tarafından okunmuş, profesyonel olarak kaydedilmiş çeşitli cümlelerden oluşmaktadır [3] (Arctic bdt, slt, clb, rms). Bu bölümün girişinde anlatıldığı üzere elimizdeki her sese Fourier analizi ve spektrum analizi yapılmalıdır. Analizin frekans uzayında gerçekleştirilmesinin arkasındaki motivasyon, formant olarak bilinen, sese ait karakteristik özelliklerin her iki cinsiyette genellikle farklılıklar göstermesinden yararlanarak sesler üzerinden cinsiyete dair ayrımının yapılabilmesini sağlamaktır.[4] Frekans analizlerinde kullanılan ses dosyalarının hepsinde aynı cümleler, aynı sıralarla bu 4 farklı kişi tarafından okunmaktadır.



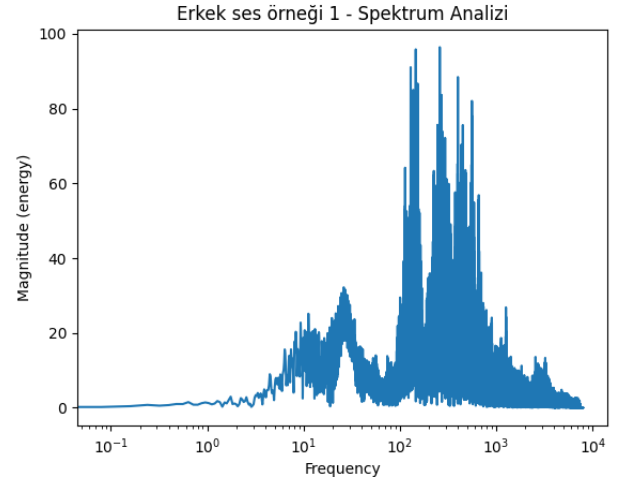
Şekil 1. Kadın 1 ses örneği Fourier dönüşümü



Şekil 2. Kadın 1 ses örneği spektrum analizi

Veritabanından ses dosyaları ikiye ayrılarak iki gruba ayrıldı; female1, male1 ve female2, male2. Bu ses kayıtlarında 4 farklı seslendirmenin okudukları cümleler aynı olduğu için ses karakterlerinin anlaşılması daha kolaylaşmıştır. Amaç bu seslerdeki temel farklılıkların bulunup, bu farklılıkları açıkça belli edecek filtreler tasarlamaktır. Kadın sesini erkek sesinden ayıran belirli özellikler vardır. Temel frekans, ve normal konuşma aralığının frekans değerleri gibi bazı özellikler bu ayrımı yapmakta kullanıldı.[5]

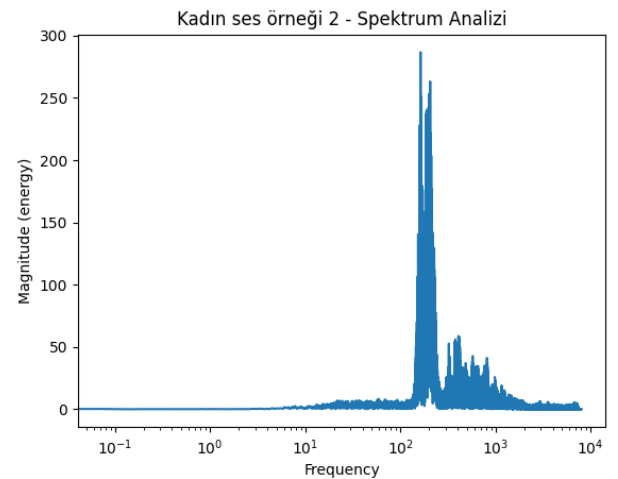
Şekil 2'de görüldüğü üzere ilk kadın ses örneğinde (female1\_comb.wav) açık olarak göz ile görülen 4 tane formant bulunmaktadır. İlk formant temel frekansın kendisi olup kabaca 190 Hz'e denk gelmektedir.



Şekil 3. Erkek 1 ses örneği spektrum analizi

Şekil 3'te (male1\_comb.wav) formantların female1.wav'daki formantlardan daha düzensiz genliklere sahip olduğu görülmektedir. İlk ortaya çıkan formant 138 Hz'te olup sesin temel frekansı değildir. 50 Hz civarında gözükten küçük tepe okuyan kişinin 't', 'p', 'f' gibi harfleri okurken mikrofonu üflemesinden dolayı kaynaklanmış olup bir problem teşkil etmemektedir.

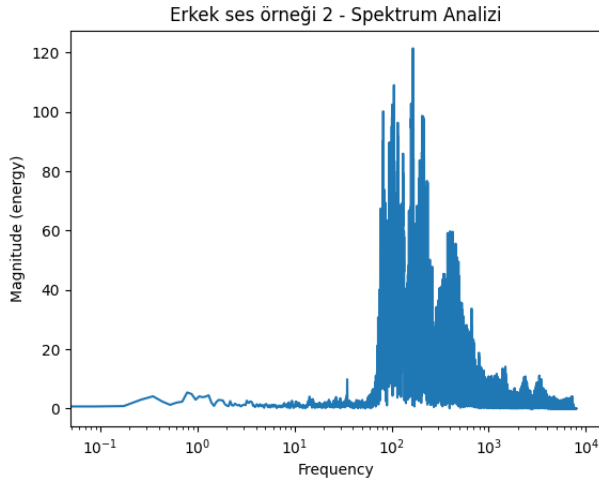
**Çıkarım:** Sadece bu iki spektrum analizinden çıkan yaklaşım erkeklerin temel konuşma frekansları altında bir pes tonları olduğunu gösterirken, kadınların formant ilerleyişlerinin çok daha düzgün olup temel konuşma frekanslarından daha tiz tonlara doğru genlikleri azalarak ilerlediğini göstermektedir. Bu çıkarımın daha net gözlemlenmesi için iki tane daha, farklı erkek ve kadın ses örnekleri üzerinde ölçümler yapıldı.



Şekil 4. Kadın 2 ses örneği spektrum analizi

Şekil 4'te (female2\_comb.wav) formantların yapısının önceki gözlemlenen grafiklerden daha yalın olduğu görüldü. Bu durum, örnekleri alınan bu kişinin sesindeki temel frekansın genlik ağırlığının diğer formantlardan çok daha fazla olmasından kaynaklanmaktadır. Yapılan ölçümden bu formantın 193 Hz'te bulunduğu

görülmektedir. Sesin karakterini yaratan yüksek frekanslardaki formantların bu kadar yüksek genlikte bulunmamasından ötürü yüksek geçiren filtre tasarımında bir kesim noktası seçmekte güçlük çıkmıştır.



Şekil 5. Erkek 2 ses örneği spektrum analizi

Şekil 5'te (male2\_comb.wav) önceki erkek örneğine benzer bir yapı gözlemlemekteyiz fakat bu örneğin genel konuşma aralığı önceki erkek sesine kıyasla daha pes tonlarda olup, ilk formant 103 Hz'te bulunmaktadır.

**Sonuç:** İlk iki örnekten sonraki çıkarımda bir problem bulunmamaktadır, sadece farklı karakterlere sahip iki tane daha ses ile önceki tasdiklenmiş olup erkek ve kadın seslerindeki formant dizilimleri hakkında fikir sahibi olunmuştur.

## 2.2. Filtre Oluşturma

Önceki çıkarımlarda da görüldüğü üzere erkek ve kadın sesi arasında bazı yapısal farklılıklar bulunmaktadır. Formant yapı farklılıkları bir yana, genel konuşma seslerinin frekans aralıkları farklı olup, bu farklılıklara dayanılarak bir ayırım yapılabileceği düşünüldü.

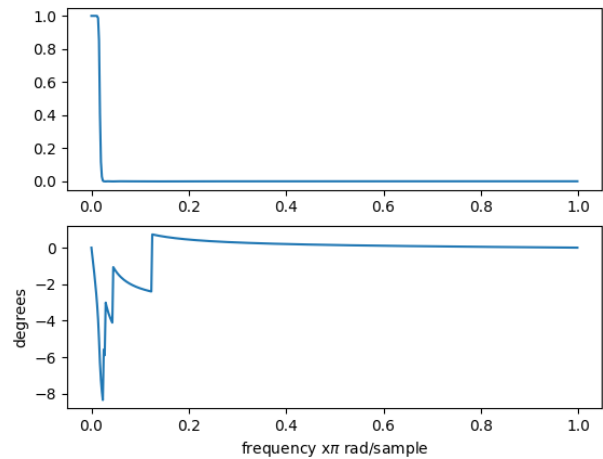
*Erkek sesleri için;* erkek sesini kadın sesinden ayıran en büyük etmen, temel formanttan bir önceki düşük frekanslı formanttır.[6] Bu formantın bulunduğu frekans ikinci erkek sesinde görüldüğü üzere 103 Hz kadar inebilirken kadın seslerinde en düşük görülen formantın olduğu frekans 193 Hz'tir. Bu yüzden iki erkek örneğini de içinde barındıracak bir Alçak Geçiren Filtre tasarlanması uygun görüldü. Geçmesi gereken kısmın seçilebilmesi için  $f_{PB}$  parametresi gereğince ayarlanmıştır.

**Alçak Geçiren Filtre,** pyFDA ortamında tasarlanan filtrenin parametreleri Şekil 6'daki gibidir. Tasarlanan filtre daha dik bir durdurma bandı olduğundan dolayı Chebyshev Tip-2 türünde filtreler kullanılmıştır. (pyFDA ekran görüntüsünde gözükten değerler 0,  $1\pi$  aralığındadır) pyFDA ortamında tasarlanan filtrenin parametreleri .mat

dosyası olarak dışa aktarılmış olup, Python'da oluşturulmuştur.

Şekil 6. pyFDA ortamında tasarlanan AGF'nin parametreleri

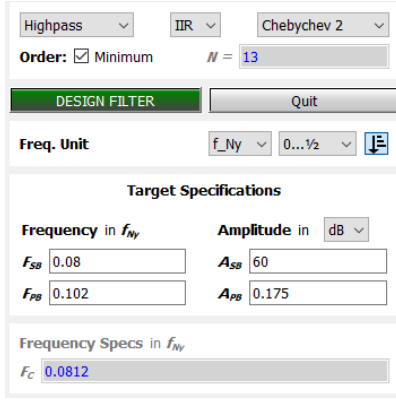
Erkek Sesi İçin Tasarlanan Alçak Geçiren Filtre,  $F_c = 0.0243\pi$



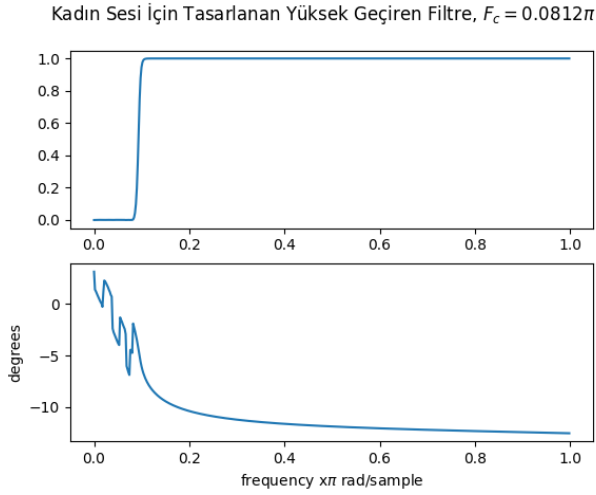
Şekil 7. Tasarlanan AGF'nin Python ile oluşturulması

*Kadın sesleri için;* kadın sesinin yüksek frekanslarda da birçok formantı bulunur. Bu frekanslardaki formantlar erkeklerde de bulunmasına rağmen genlik değerleri kadın seslerindekiyle nazaran azdır. Yapılan ölçümlerde 600 ile 1000 Hz arasındaki formantların genlik değerleri erkekler için tasarlanan AGF nin geçirdiği aralıktaki değerlerden daha yukarıdadır. Bu yüzden iki kadın örneğini de içinde barındıracak (Kadın 2 örneği incelenirken belirtildiği üzere kişinin ses karakteristiğinden ötürü bir takım komplikasyonlar ortaya çıkmıştır. Bu sorunlar uygulama bölümünde detaylı olarak incelenmiştir.) Geçmesi gereken kısmın seçilebilmesi için  $f_{SB}$  parametresi gereğince ayarlanmıştır.

**Yüksek Geçiren Filtre,** pyFDA ortamında tasarlanan filtrenin parametreleri Şekil 8'deki gibidir. Tasarlanan filtre daha dik bir durdurma bandı olduğundan dolayı Chebyshev Tip-2 türünde filtreler kullanılmıştır. (pyFDA ekran görüntüsünde gözükten değerler 0,  $1\pi$  aralığındadır) pyFDA ortamında tasarlanan filtrenin parametreleri .mat dosyası olarak dışa aktarılmış olup, Python'da oluşturulmuştur.



Şekil 8. pyFDA ortamında tasarlanan YGF'nin parametreleri

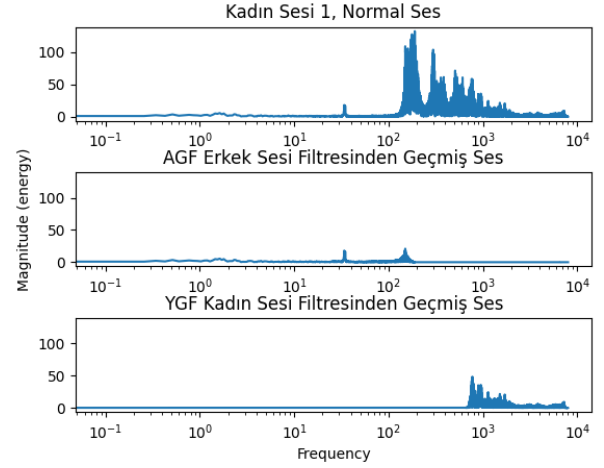


Şekil 9. Tasarlanan AGF'nin Python ile oluşturulması

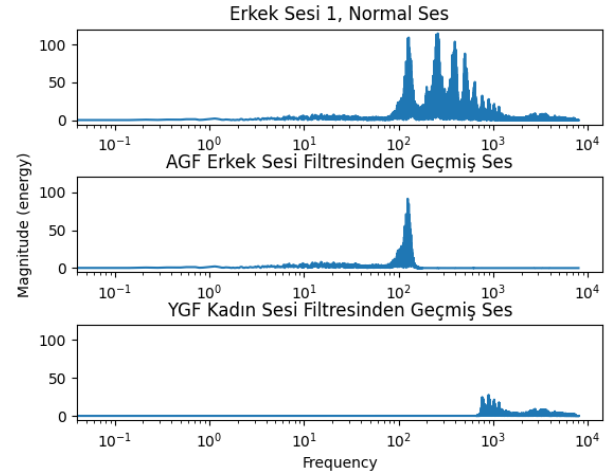
### 2.3. Uygulama

Aynı 4 kişinin başka ve birbirinden farklı rastgele cümleler söylediği 4 tane daha ses dosyası (female1\_test.wav, male1\_test.wav, female2\_test.wav, male2\_test.wav) oluşturulduktan sonra, önceki aşamada oluşturulan filtrelerden geçirilip spektrum analizleri yapılmıştır.

Şekil 10'daki durumda ana ses kaynağının Kadın 1 olduğunun bilinmediğini varsayarsak, tasarlanan iki filtrenin çıkışlarının verdiği değerlerden cinsiyet ayrımı yapılabilmektedir. AGF'den neredeyse hiçbir ses geçmez iken, YGF'nin çıkışında oluşan sesin ortalama genliği AGF'den çıkana kıyasla daha yüksektir, bu da Frekans Analizi kısmında vardığımız sonucu doğrulamaktadır. Bu yüzden incelenen ses bir kadın sesidir.

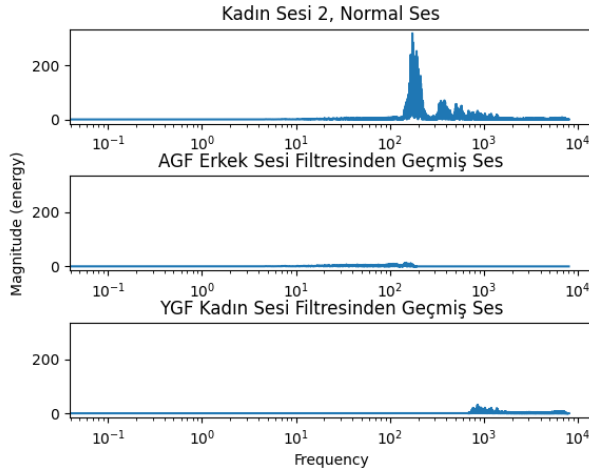


Şekil 10. Kadın 1, normal ve filtrelenmiş hallerin spektrum analizleri



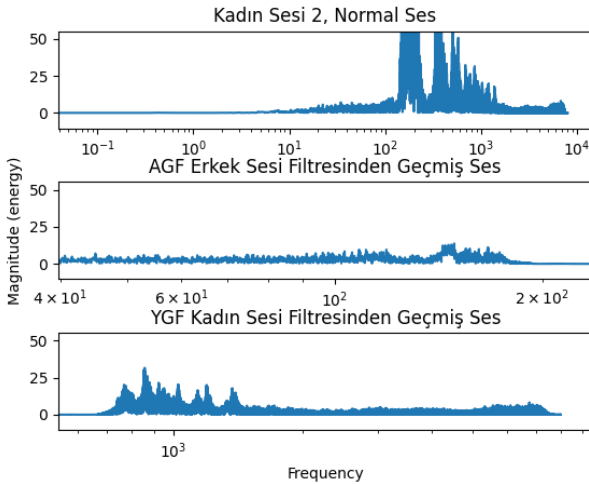
Şekil 11. Erkek 1, normal ve filtrelenmiş hallerin spektrum analizleri

Şekil 11'deki durumda ana ses kaynağının Erkek 1 olduğunun bilinmediğini varsayarsak, tasarlanan iki filtrenin çıkışlarının verdiği değerlerden cinsiyet ayrımı yapılabilmektedir. AGF'den erkek sesindeki ilk pes tondaki formant geçerken, YGF'den geçen sesin genliği diğer filtrenin çıkışına kıyasla çok küçük kalmaktadır, bu da Frekans Analizi kısmında vardığımız sonucu doğrulamaktadır. Bu yüzden incelenen ses bir erkek sesidir.

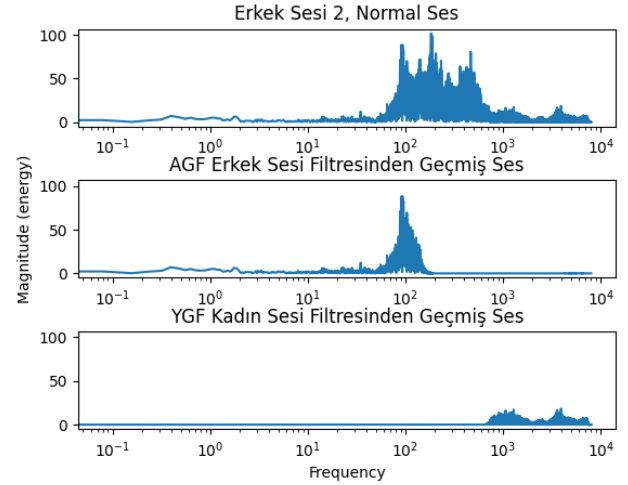


**Şekil 12.** Kadın 2, normal ve filtrelenmiş hallerin spektrum analizleri

Şekil 12'deki durumda ana ses kaynağının Kadın 2 olduğunun bilinmediğini varsayarsak, tasarlanan iki filtrenin çıkışlarının verdiği değerlerden cinsiyet ayrımı yapılabilmektedir. Burada önceki kadın sesinden farklı olarak ana formantın boyutunun çok büyük olup yan formantlarının boyutları daha küçüktür. AGF'den neredeyse hiçbir ses geçmezken, YGF'nin çıkışında oluşan sesin ortalama genliği şu anki halinde bile AGF'den çıkana kıyasla yüksektir (*bkz.* Şekil 13, y eksenini tüm grafiklerde aynı ölçeğe getirilirse), burada sadece bu farka bakılmayıp AGF'den neredeyse hiç bir sesin geçmemesi de bu sesin bir erkek olmayacağını anlatır. Bu da Frekans Analizi kısmında vardığımız sonuca uymaktadır. İncelenen sesin bir kadın sesi olduğu söylenebilmesine rağmen iki filtrenin çıkışları arasındaki küçük farktan dolayı tamamen emin olunamamaktadır.



**Şekil 13.** Kadın 2, spektrum analizlerinin yakınlaştırılmış halleri



**Şekil 14.** Erkek 2, normal ve filtrelenmiş hallerin spektrum analizleri

Şekil 14'deki durumda ana ses kaynağının Erkek 2 olduğunun bilinmediğini varsayarsak, tasarlanan iki filtrenin çıkışlarının verdiği değerlerden cinsiyet ayrımı yapılabilmektedir. Aynı Erkek 1 örneğinde olduğu gibi AGF'den erkek sesindeki ilk pes tondaki formant geçerken, YGF'den geçen sesin genliği diğer filtrenin çıkışına kıyasla çok küçük kalmaktadır, bu da Frekans Analizi kısmında vardığımız sonucu doğrulamaktadır. Bu yüzden incelenen ses bir erkek sesidir.

### 3. Sonuç ve Yorum

Bu proje sonucunda, bir ses işaretinin Fourier ve spektrum çıktılarının nasıl yorumlanacağı hakkında fikir sahibi olundu. Bu analizlere göre kadın ve erkek sesleri özellikle formant denilen karakteristik özellikleriyle birbirlerinden ayrılabilir. Bu analize göre erkek sesi temel konuşma frekansı altında daha pes bir ton barındırırken, kadın sesi temel frekansının yanında daha tiz alt tonları barındırmaktadır. Bu yorumlamanın getirisi olarak bu ses işaretlerine uygun bir filtre tasarlarken formantlarının ses işaretlerini kadın ve erkek olarak iki sınıfa ayıracak şekilde filtrelenmesine özen gösterilmelidir. Filtrelemenin Alçak Geçiren Filtre ve Yüksek Geçiren Filtre ikilişiyle gerçekleştirilmesi tercih edilirken kadın sesinin içerdiği tiz bileşenler sebebiyle YGF ile gözlemleneceği ancak AGF'nin kadına ait formantları barındıramayacağı açıktır. Ayrıca erkek sesinde yüksek frekanslarda bulunan formantların YGF'den geçmesine rağmen erkeklerin AGF'den geçmiş hallerinin genlikleri YGF'den geçen küçük kısımdakinden daha fazladır.

İnsan konuşma sesine ait frekanslar cinsiyetlere göre bariz bir farklılık teşkil etmemektedir.[7] Bu durum eldeki her çeşit veri setine aynı başarıyla cevap verilmesini zorlaştırmaktadır. Bu durumun bir örneği Şekil 12'deki kadının sesinin tiz bileşenlerinin genliklerinin diğer

kadın ses örneğindeki bu bileşenlerin genliklerinden farklılığı ile verilebilir. Genlik değerlerindeki bu düşüş aynı parametrelere sahip YGF'lerin her iki örneğe uygulanmasını zorlaştırırken, bu ses örneği için YGF'nin çıkışında belirgin frekans bileşenleri gözlemlenememiştir.

Temel olarak cinsiyetlere ait frekans bileşenlerinin benzerliğinden kaynaklanan sorunlar, seslerin filtreleme yoluyla iki sınıfa ayrılmasını güçleştirmektedir. Böyle bir problem için, proje kapsamında bulunmayan, makine öğrenmesi algoritmalarının kullanılması daha uygun olacaktır.

## Kaynaklar

[1] Singh, Nilu. (2015). Speaker Recognition and Fast Fourier Transform.doi:10.13140/RG.2.1.2722.0969.

[2] Mahdi, Jinan. (2019). *Frequency analyses of human voice using fast Fourier transform*. Iraqi Journal of Physics (IJP). 13. 174-181. 10.30723/ijp.v13i27.276. Available:[https://www.researchgate.net/publication/343624983\\_Frequency\\_analyses\\_of\\_human\\_voice\\_using\\_fast\\_Fourier\\_transform](https://www.researchgate.net/publication/343624983_Frequency_analyses_of_human_voice_using_fast_Fourier_transform) [Accessed: 27 May 2021]

[3] CMU\_ARCTIC speech synthesis databases. Available: [http://festvox.org/cmu\\_arctic/index.html](http://festvox.org/cmu_arctic/index.html) [Accessed: 2 June 2021]

[4] Hassam Ullah Sheikh (2013), The University of Manchester. Who Is Speaking? Male or Female. Available: [https://studentnet.cs.manchester.ac.uk/resources/library/thesis\\_abstracts/MSc13/FullText/Sheikh-HassamUllah-fulltext.pdf](https://studentnet.cs.manchester.ac.uk/resources/library/thesis_abstracts/MSc13/FullText/Sheikh-HassamUllah-fulltext.pdf) [Accessed: 2 June 2021]

[5] Adrian P. Simpson (2009). *Phonetic differences between male and female speech*. , 3(2), 621–640. doi:10.1111/j.1749-818x.2009.00125.x [Accessed: 2 June 2021]

[6] Traunmüller, Hartmut & Eriksson, Anders. (1995). *The frequency range of the voice fundamental in the speech of male and female adults*. 2. Available: [https://www2.ling.su.se/staff/hartmut/f0\\_m&f.pdf](https://www2.ling.su.se/staff/hartmut/f0_m&f.pdf) [Accessed: 2 June 2021]

[7] DPA Microphones. (2021). *FACTS ABOUT SPEECH INTELLIGIBILITY*. Available: <https://www.dpamicrophones.com/mic-university/facts-about-speech-intelligibility> [Accessed: 1 June 2021]