

ELM368 – DÖNEM PROJESİ

FİLTRELEME YÖNTEMİYLE KONUŞMANIN ARKA PLANINDAKİ STADYUM GÜRÜLTÜSÜNÜ AZALTMA

Arda Derici, Serdar Başyemenici, Umut Mehmet Erdem

200102002051, 200102002061, 200102002025

a.deric2020@gtu.edu.tr, s.basyemenici2020@gtu.edu.tr, u.erdem2020@gtu.edu.tr

ABSTRACT (ÖZET)

Proje, stadyum, spor salonu gibi gürültülü ortamlarda iki kişinin konuşmasını daha net duyulabilir hale getirmek için bir çözüm sunmaktadır. Problemin tanımı, yüksek arka plan uğultusu ve tezahürat gürültüsünün konuşma anlaşılabilirliğini büyük ölçüde azaltmasıdır. Bu projede, gürültü azaltma teknikleri kullanılarak konuşma işaretlerinin arka plan gürültüsünden ayrıştırılması hedeflenmiştir. Çalışmanın sonuçları, gürültülü ortamlarda iletişimi iyileştirmek için pratik uygulamalar sunmaktadır. Bu belge, gerçekleştirilen yöntemleri, kullanılan algoritmaları ve elde edilen sonuçları detaylı olarak incelemektedir.

ANAHTAR KELİMELELER

FIR Filtre, Gürültü Azaltma, Stadyum Tezahürat Gürültüsü, Konuşma, Spektrum Analizi

1. Giriş

Stadyumlarda maç esnasında oluşan yoğun gürültü, yan yana oturan insanların bile birbirlerini anlayamayacak hale gelmesine yol açmaktadır. Bu durum, yanlarındakiyle iletişim kuramayan seyircilerin maçtan aldıkları keyfi büyük ölçüde azaltmaktadır. Seyirciler, yanlarındaki kişilerle iletişim kurabilmek için yüksek sesle konuşmak zorunda kaldıklarından, ses tellerine zarar vermekte ve maç sonrasında rahatsızlık hissetmektedirler. Ayrıca, bu gürültü kirliliği, seyircilerin maç deneyiminden tam anlamıyla keyif almasını engelleyerek genel memnuniyeti düşürmektedir. Bu projede, spor müsabakası esnasında iki kişinin konuşmasını stadyumdaki tezahürat gürültüsünden arındırarak birbirlerini daha iyi duymalarını ve anlamalarını sağlamak ve böylece bu kişilerin seyir keyfini artırmak hedeflenmiştir. Bu amaçla, uygun filtreleme işlemi yapılmıştır. Söz konusu bu filtreleme işlemi, seyircilerin hem maç atmosferini hissetmelerini hem de rahatça iletişim kurabilmelerini aynı anda mümkün kılmaktadır.

2. Deneyler ve Analiz

Yapılacak çalışma için gerekli olan gürültülü konuşma seslerinin elde edilebilmesi amacıyla Audacity programı kullanılmıştır. Bu programdan farklı şiddette olan gürültüler ile yapay zeka kullanılarak oluşturulan 6 farklı diyalog birleştirilerek gürültülü sesler elde edilmiştir. Gürültülü seslerin hepsi 10 saniyenin altında olacak şekilde ayarlanmıştır.

Gürültülü sesler, Python da scipy kütüphanesinin wavfile modülü kullanılarak ses verisi, uzunluk, kanal sayısı ve örnekleme frekansı değerleri elde edilmiştir. Elde edilen veriler kullanılarak sesler zaman uzayında çizdirilmiş ve numpy kütüphanesinin fft modülü kullanılarak ses verilerinin frekans uzayında genlik spektrumları çizdirilmiştir. Çizdirilen genlik spektrumları incelenerek farklı kesim frekanslarına sahip filtreler elde etmek amacıyla Pyfda yazılımı üzerinden oluşturulmuştur. Oluşturulan filtreler, eğitim için kullanılacak gürültülü sesler üzerinde scipy kütüphanesinin lfilter modülü ile uygulanmış ve Ipython kütüphanesinin display ve Audio modülleri wav uzantılı ses dosyasına dönüştürülerek test edilmiştir. Eğitim için kullanılan gürültülü seslere en uygun filtreler belirlendikten sonra test amacıyla kullanılacak sesler için sırasıyla hangi filtrenin kullanılması gerektiğine karar verilmiştir ve uygulanmıştır.

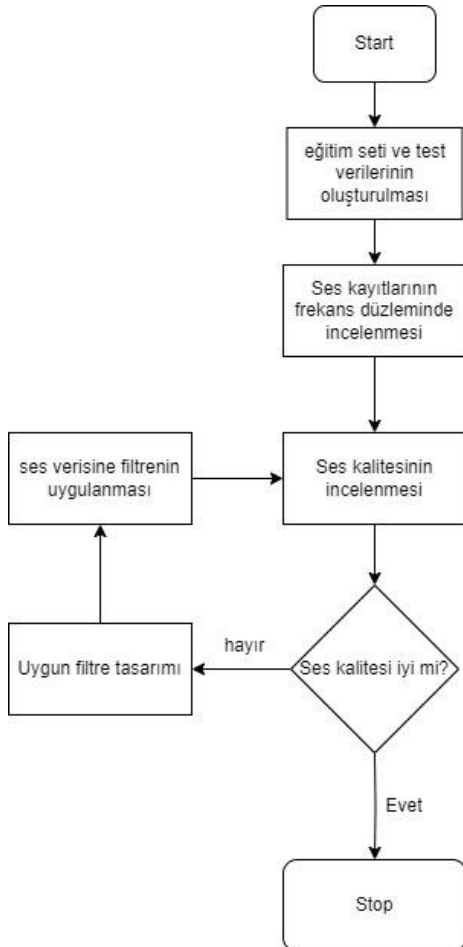
2.1. Problemin Çözümü

Belirlenen problemin çözümü için öncelikle literatür taraması yapılmış; akademik makaleler, tezler, projeler ve dijital kaynaklar incelenmiştir. Bu incelemeler sonucunda birkaç çözüm yöntemi bulunmuş olup, bu düzeyde uygulanacak çözüm yöntemi olarak filtreleme yöntemi seçilmiştir. Proje aşağıdaki adımlarla ilerletilmiştir.

- 1. Veri Toplama:** Arka planda stadyum gürültüsü olan iki kişinin konuşmasının olduğu ses verileri alınmıştır.
- 2. Veri İşleme:** Alınan bu ses verileri Python'da işlenmiştir.
- 3. Veri Görselleştirme:** Söz konusu bu ses verileri x-ekseni örnek olacak şekilde ayrıklı zamanda çizdirilip dinlenmiştir.

4. **Fourier Dönüşümü:** Spektrum analizi yapabilmek amacıyla ses verilerinin Fourier dönüşümü alınmıştır.
5. **Spektrum Analizi:** Gürültü azaltmak amacıyla filtre tasarımı için ses verilerinin spektrum analizi yapılmıştır.
6. **Frekans İncelemesi:** Gürültü ve konuşma sesinin spektrumları incelenmiş, insan seslerinin 90 Hz ile 255 Hz arasında yoğunlaştığı araştırılarak ^[1] belirlendiğinden bu frekansları ayırt etmek için uygun filtreleme yöntemleri seçilmiştir. $\omega = 2\pi \frac{f}{f_s}$ eşitliği kullanılarak ω -eksenindeki karşılık gelen değerler hesaplanmıştır. Kullanılan tüm ses verilerinin örnekleme frekansı $f_s = 44100$ Hz değerindedir.
7. **Filtreleme Uygulaması:** Belirlenen filtreleme yöntemleri ve hesaplanan değerler uygulanarak mümkün olduğunca gürültü azaltılmış ve konuşmaların anlaşılabilirliği artırılmıştır.
8. **Sonuçların Değerlendirilmesi:** Sonuçlar analiz edilerek elde edilen verimlilik değerlendirilmiş, proje için gerekli yorumlar ve çıkarımlar yapılmıştır.

Bu adımların ve projenin algoritmasının akış diyagramı Şekil 1'de gösterilmiştir.

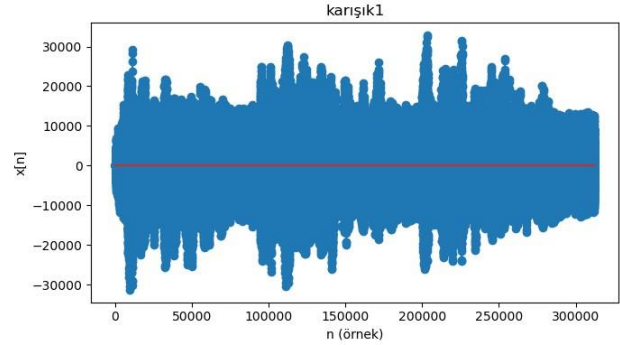


Şekil 1 – Akış diyagramı.

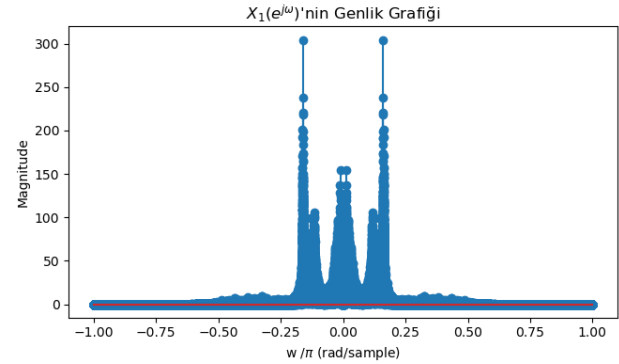
2.2. Grafik, Tablo ve Şekiller

2.2.1. Eğitim Verilerinin Grafikleri

Proje kapsamında eğitim için 3 adet veri işlenmiştir. Eğitim için olan ilk ses verisinin kaydının grafiği Şekil 2, genlik spektrumu ise Şekil 3'tedir.

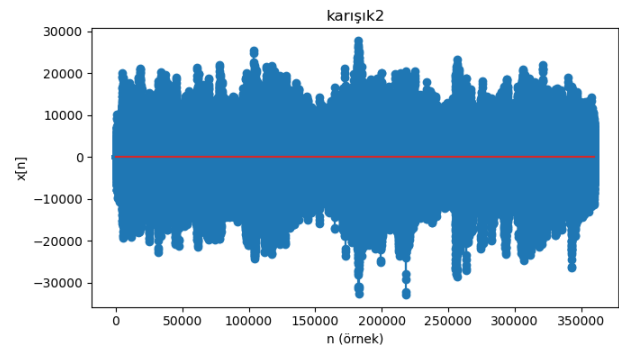


Şekil 2 – İlk eğitim verisinin ses kaydı grafiği.

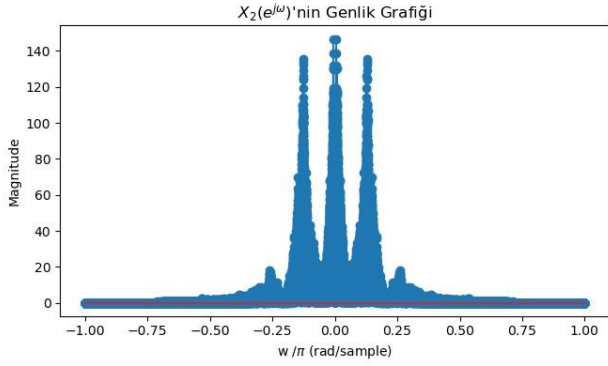


Şekil 3 – İlk eğitim verisinin genlik spektrumu.

Eğitim için olan ikinci ses verisinin kaydının grafiği Şekil 4, genlik spektrumu ise Şekil 5'tedir.

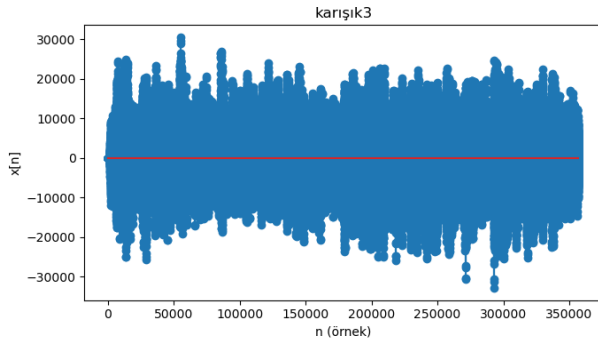


Şekil 4 – İkinci eğitim verisinin ses kaydı grafiği.

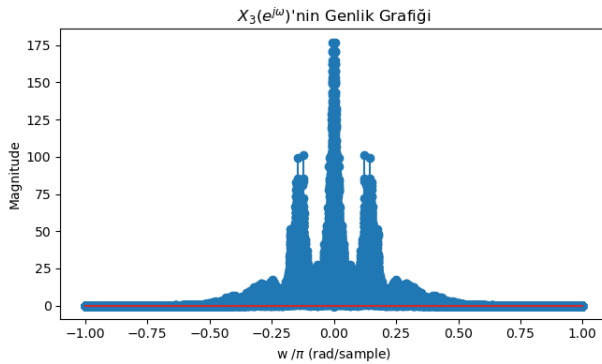


Şekil 5 – İkinci eğitim verisinin genlik spektrumu.

Eğitim için olan üçüncü ses verisinin kaydının grafiği Şekil 6, genlik spektrumu ise Şekil 7’dir.



Şekil 6 – Üçüncü eğitim verisinin ses kaydı grafiği.



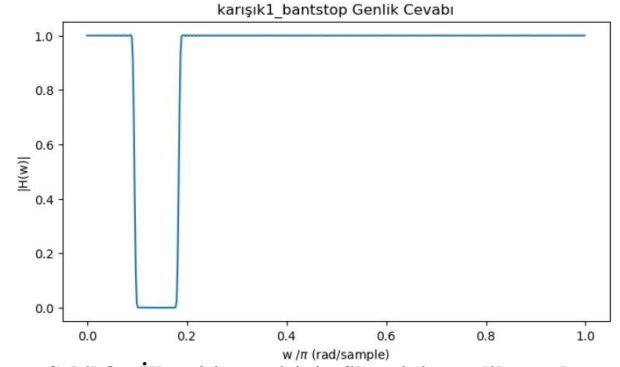
Şekil 7 – Üçüncü eğitim verisinin genlik spektrumu.

Eğitim için kullanılan ses verilerinin farklı konuşmalar ve farklı gürültüler kullanılmasına rağmen benzer genlik spektrumlarına sahip olduğu gözlemlenmiştir. Yüksek genlik değerlerinin birbirlerine yakın frekanslarda olduğu gözlemlendiğinden gürültü ve konuşmanın birbirine karıştığı düşünülmektedir.

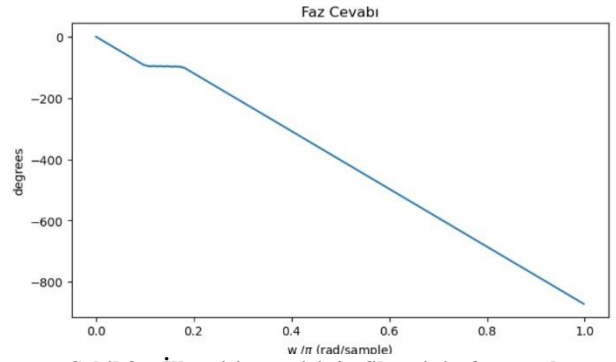
2.2.2. Filtre Tasarımı

İlk eğitim verisi için birçok farklı filtre denenmiştir. İnsan sesinin bulunmadığı frekanslar gürültü olarak düşünülüp o frekans değerlerinde bant-stop filtresi uygulanmıştır. Bu yapılan filtrelemenin en anlaşılır konuşmayı verdiği

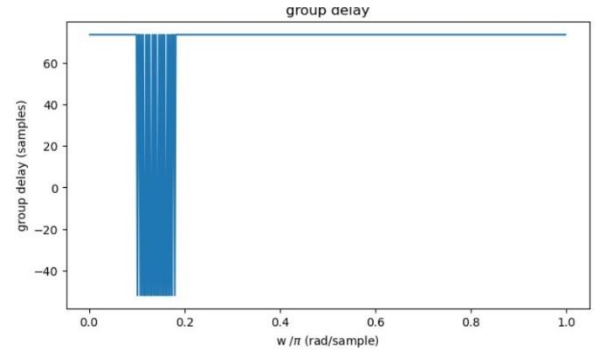
görülmüştür. Filtrenin genlik cevabı, faz cevabı ve grup gecikmesi grafikleri sırasıyla Şekil 8, 9 ve 10’da görülebilir.



Şekil 8 – İlk eğitim verisinin filtresinin genlik cevabı.

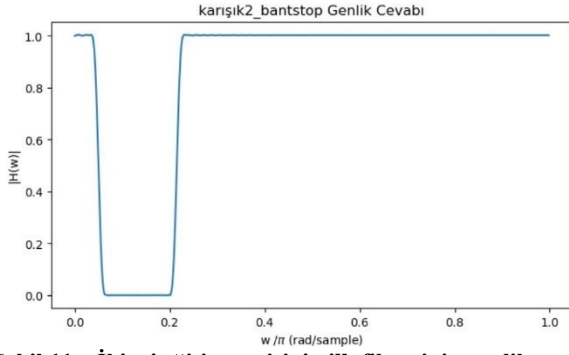


Şekil 9 – İlk eğitim verisinin filtresinin faz cevabı.

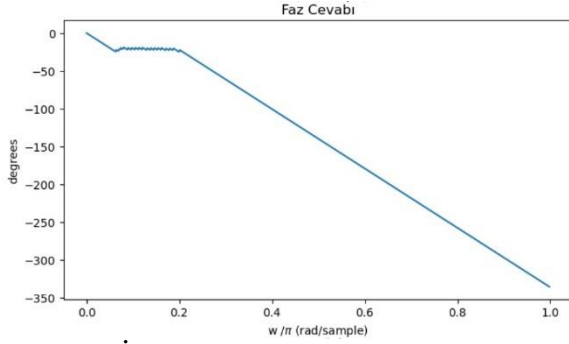


Şekil 10 – İlk eğitim verisinin filtresinin grup gecikmesi.

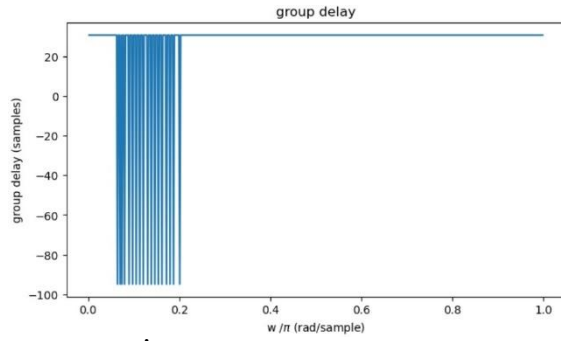
İkinci eğitim verisi için filtre oluşturulurken, ilk eğitim verisinden yararlanılmıştır. Genlik cevabında gürültünün bulunduğu frekans bölgeleri belirlenmiş ve bu bölgeler iki ayrı bant-stop filtresi ile bastırılarak konuşmanın en net hali elde edilmiştir. İlk bant-stop filtresinin genlik cevabı, faz cevabı ve grup gecikmesi grafikleri sırasıyla Şekil 11, 12 ve 13’te görülebilir.



Şekil 11 – İkinci eğitim verisinin ilk filtresinin genlik cevabı.

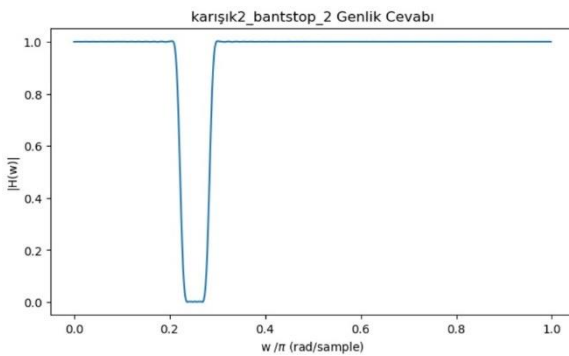


Şekil 12 – İkinci eğitim verisinin ilk filtresinin faz cevabı.

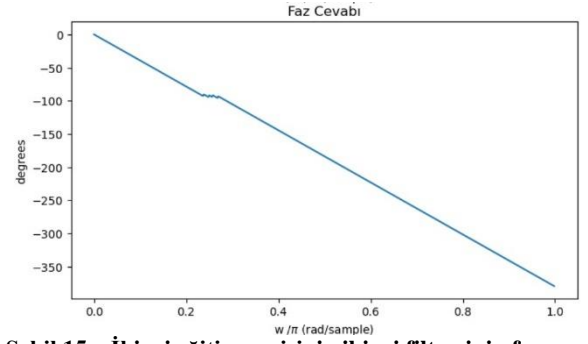


Şekil 13 – İkinci eğitim verisinin ilk filtresinin grup gecikmesi.

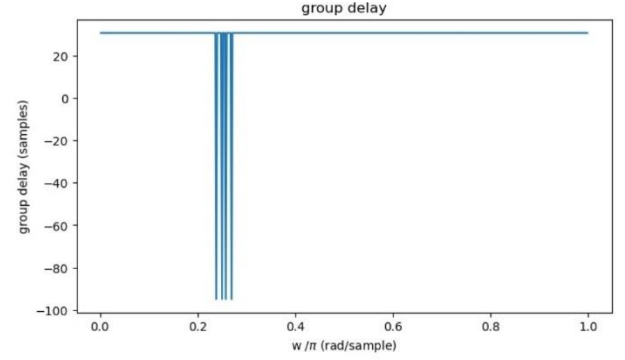
İkinci eğitim verisi için tasarlanan ikinci bant-stop filtresinin genlik cevabı, faz cevabı ve grup gecikmesi grafikleri sırasıyla Şekil 14, 15 ve 16'da görülebilir.



Şekil 14 – İkinci eğitim verisinin ikinci filtresinin genlik cevabı.

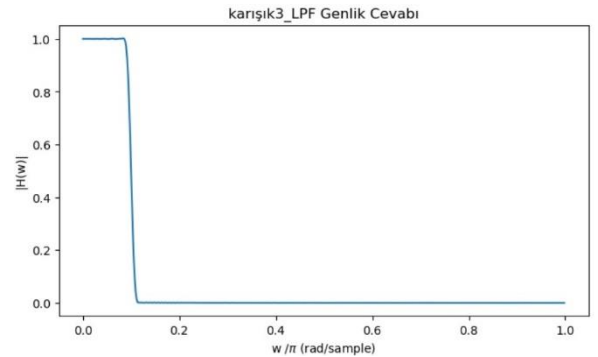


Şekil 15 – İkinci eğitim verisinin ikinci filtresinin faz cevabı.

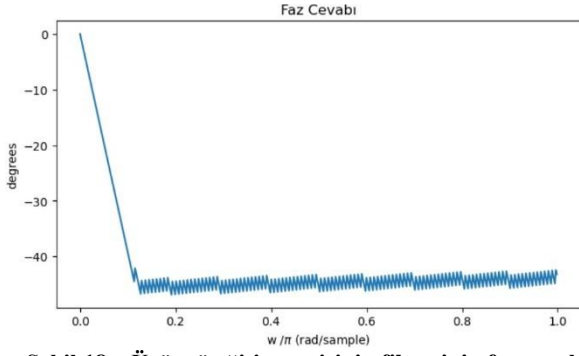


Şekil 16 – İkinci eğitim verisinin ikinci filtresinin grup gecikmesi.

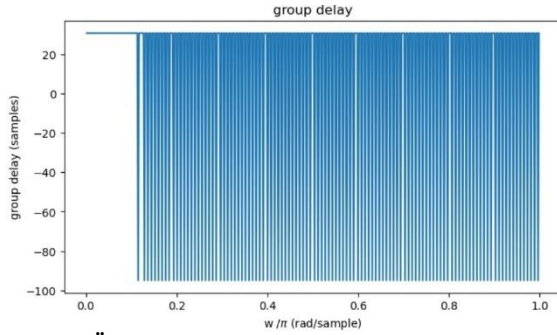
Üçüncü eğitim verisi için ilk iki eğitim verisinde kullanıldığı gibi bant-stop filtresi kullanıldığında gürültünün yeterince ayıklanmadığı fark edilmiştir. Bu nedenle insan sesinin en yoğun bulunduğu frekans aralığı için bir alçak geçiren filtre uygulandığında gürültüden yüksek oranda arındırılmıştır. Ancak ses kalitesinden de kayıp verilmiştir. Üçüncü eğitim verisi için tasarlanan alçak geçiren filtrenin genlik cevabı, faz cevabı ve grup gecikmesi grafikleri sırasıyla Şekil 17, 18 ve 19'da görülebilir.



Şekil 17 – Üçüncü eğitim verisinin filtresinin genlik cevabı.



Şekil 18 – Üçüncü eğitim verisinin filtresinin faz cevabı.



Şekil 19 – Üçüncü eğitim verisinin filtresinin grup gecikmesi.

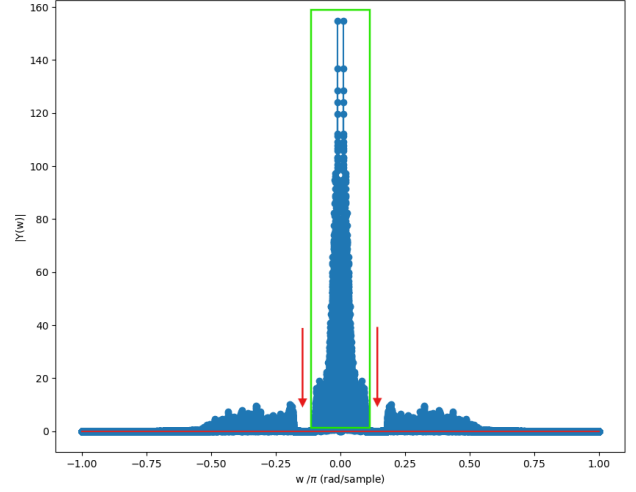
Bu filtre tasarımlarında Finite Impulse Response (FIR) filtreler kullanılmıştır. Bunun birkaç sebebi vardır. Bunlardan biri doğrusal faz karakteristiğine sahip olmasıdır. Bu durum filtrenin sabit grup gecikmesine sahip olmasını sağlar, yani tüm frekans bileşenleri aynı miktarda gecikir. Filtre çıkışındaki işaretin faz bozulması olmadan elde edilmesine olanak tanır. Bunun yanı sıra, FIR filtreler tasarım açısından büyük esneklik sunarlar. İstenilen frekans yanıtını elde etmek için kolayca şekillendirilebilirler.

FIR filtre türü olarak Hamming penceresi tercih edilmiştir. Bunun bir sebebi, Hamming penceresinin yan lobları baskılama özelliğinden dolayı yan lobların genliğini oldukça düşük seviyede tutmasıdır. Bunun sonucunda, özellikle düşük seviyeli yan loblarda istenmeyen frekans bileşenlerinin (aliasing etkisi gibi) baskılanmasına yardımcı olarak parazitin azaltılmasına katkı sağlar. Bir diğer sebebi ise, Hamming penceresi ana lobun genişliğini orta seviyede tutarken, ana lobdaki genlik kaybını da düşük seviyede tutar. Bu, sinyalin ana bileşenlerinin daha doğru bir şekilde korunmasını sağlar.

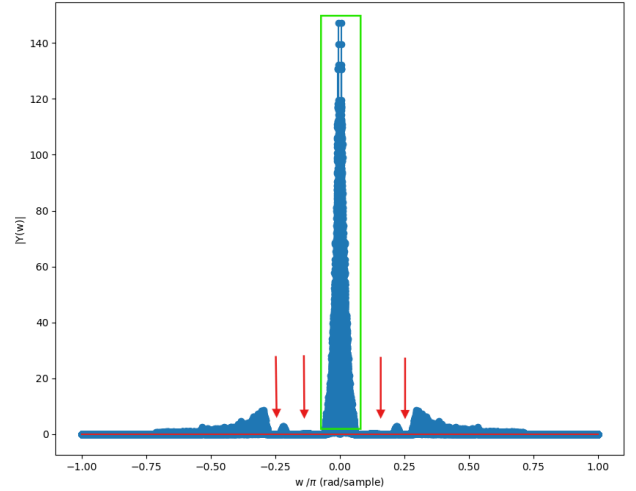
2.2.3. Filtrenin Eğitim Verilerine Uygulanması

Eğitim verilerinin genlik spektrum grafiklerine bakarak frekans ekseninde 0 (radyan/sample) etrafında insan seslerinin frekanslarının yoğunlukta olduğu, diğer yüksek genlikli olan değerlerin yoğun olduğu frekans bölgelerinin ise gürültü frekans bölgeleri olduğu düşünülmüştür. Böyle düşünülmesinin sebebi, literatür araştırması yapıldığında ^[1]

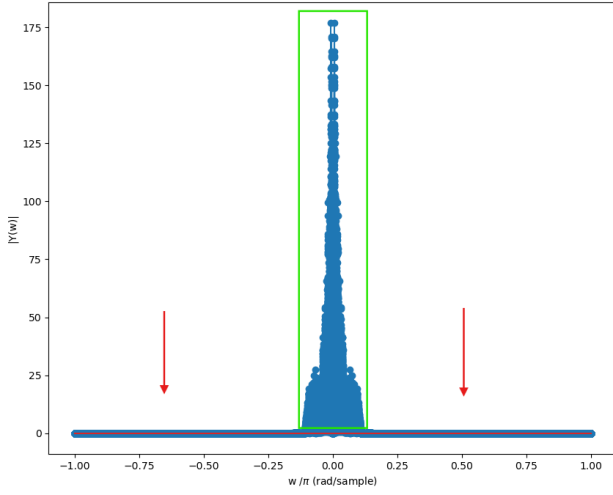
insan sesinin temel frekans aralığının 90 Hz ile 255 Hz arasında yoğunlaşmasındandır. Bunu tespit etmek amacıyla, grafiklerden belirlediğimiz yoğun frekans bölgeleri için filtreler tasarlanmış ve eğitim verilerine uygulanmıştır. Filtrelenmiş seslerin grafikleri Şekil 20, Şekil 21 ve Şekil 22’deki gibidir.



Şekil 20 – Birinci eğitim verisine birinci filtrenin uygulanması sonucunda çıkışta görülen genlik cevabı.



Şekil 21 – İkinci eğitim verisine ikinci eğitim verisi için oluşturulan filtrelerin uygulanması sonucunda çıkışta görülen genlik cevabı.



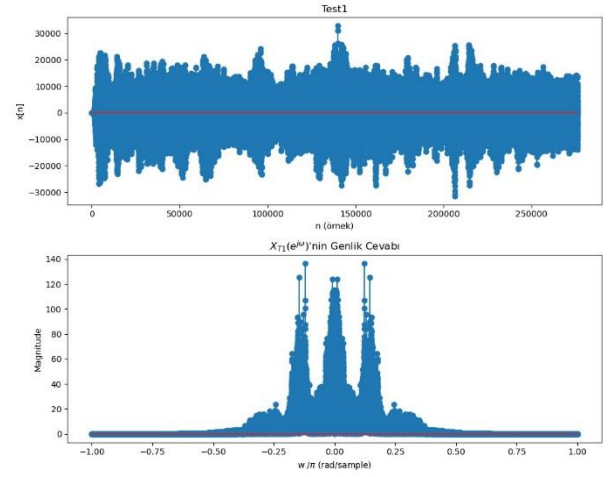
Şekil 22 – Üçüncü eğitim verisine üçüncü filtrenin uygulanması sonucunda çıkışta görülen genlik cevabı.

Şekil 20, Şekil 21, Şekil 22’deki spektrumlar incelenip filtrelenen sesler dinlendiğinde eğitim verilerinin genlik spektrumlarında yüksek frekans aralığında genliği yüksek olan bölgelerin gürültü frekans bölgeleri olduğu görülmüştür. Diğer eğitim setleri için de aynı yöntem uygulanmış ve beklenen sonuçlar elde edilmiştir. Ancak eğitim seslerinin tezahürat gürültüsünden tamamen kurtarılamadığı filtrelenmiş sesler dinlenince fark edilmiştir. Bu sorundan yola çıkılarak literatür taraması yapıldığında Wiener filtresi öğrenilmiştir. Bu filtre, optimum bir gürültü azaltma filtresi olarak bilinir ve işaret işleme alanında yaygın olarak kullanılır.^[2] Wiener filtresi gürültü ve konuşma verileri hakkında yeterli bir bilgiye sahip olması durumunda filtrelenmiş sinyal ile orijinal sinyal arasındaki ortalama kare hatayı (mean square error - MSE) minimize eder. Bu, filtreden geçirilen sinyal ile orijinal sinyal arasındaki farkın karesinin ortalamasını en küçük yapmayı amaçlar.

Sorunu çözebilecek başka bir yöntem olarak ise adaptif filtreleme yöntemi bulunmuştur.^[3] Bu filtreler, çevrelerin işaret ve gürültü koşullarına dinamik olarak uyum sağlayarak çalışan filtrelerdir. Bu tür filtreler, filtre katsayılarını sürekli olarak ayarlayarak işaret işlemede daha etkili bir performans sergilerler.

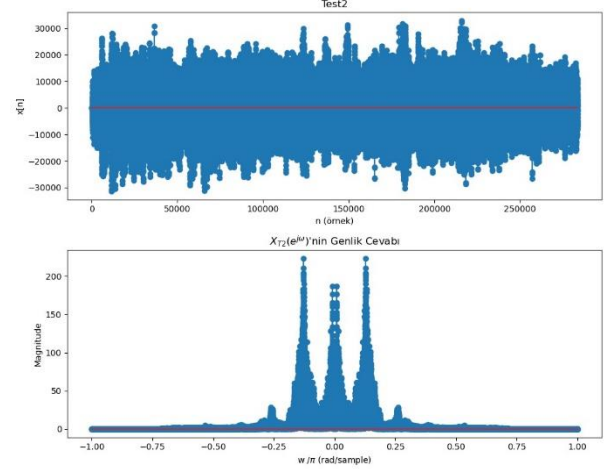
2.2.4. Test Verilerinin Grafikleri

Önceki bölümlerde eğitim için 3 adet veri işlenmiş, bu eğitim verilerini tezahürat gürültüsünden kurtarmak için filtreler tasarlanmıştır. Bu aşamada da 3 adet test verisi, Python üzerinden işlenmiş ve fft() ile genlik spektrumları çizdirilmiştir. Test için olan ilk ses verisinin kaydının grafiği ve genlik spektrumu Şekil 23’tedir.



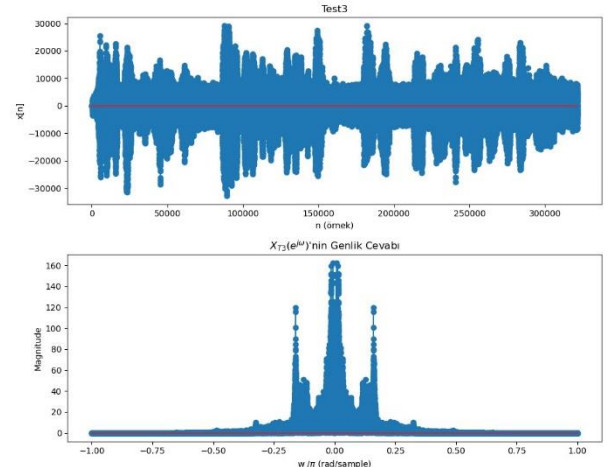
Şekil 23 – İlk test verisinin ses kaydı grafiği ve genlik spektrumu.

İkinci testin ses verisinin kaydının grafiği ve genlik spektrumu Şekil 24’tedir.



Şekil 24 – İkinci test verisinin ses kaydı grafiği ve genlik spektrumu.

Üçüncü testin ses verisinin kaydının grafiği ve genlik spektrumu Şekil 25’tedir.

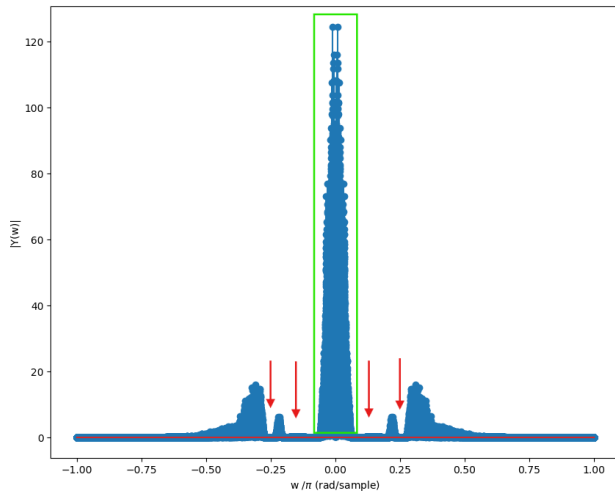


Şekil 25 – Üçüncü test verisinin ses kaydı grafiği ve genlik spektrumu.

İnsan sesinin frekansları genellikle benzer frekanslarda (90 Hz ile 255 Hz arasında) olduğu için Şekil 23, 24 ve 25'teki testlerin genlik spektrumları ile Şekil 3, 5 ve 7'deki eğitim verilerinin genlik spektrumlarının benzer oldukları görülmüştür. Bu nedenle tasarlanan filtrelerin uygun olacağı düşünülmektedir.

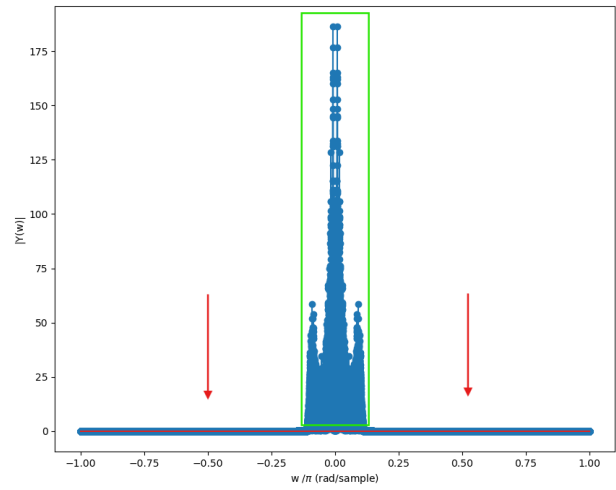
2.2.5. Filtrenin Test Verilerine Uygulanması

Şekil 26'da birinci test verisine 3 ayrı eğitim verisinde kullanılan filtreler uygulanmıştır. Aralarında en anlaşılabilir olan ses, ikinci eğitim verisinde kullanılan filtreler uygulanarak elde edilmiştir. Buradaki hedef, sesin kalitesinden en az şekilde kayıp vererek gürültüyü olabildiğince azaltmaktır.



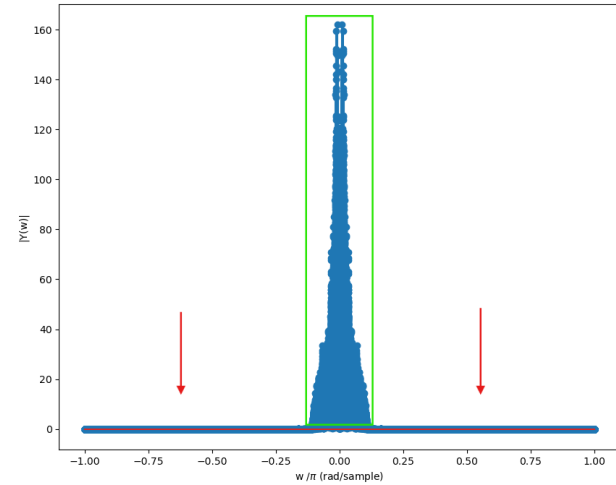
Şekil 26 – Birinci test verisine ikinci eğitim verisi için oluşturulan filtrelerin uygulanması sonucunda çıkışta görülen genlik cevabı.

Şekil 27'de ikinci test verisine 3 ayrı eğitim verisinde kullanılan filtreler uygulanmıştır. İkinci test verisinde gürültü oldukça fazla olduğundan gürültünün rahatsız edici frekanslarından kurtulmak hedeflenmiştir. Bunu en iyi gerçekleştiren filtre, üçüncü eğitim verisine uygulanan alçak geçiren filtre olduğu belirlenmiştir. Yapılan filtreleme sonucunda ses beklenenden daha kalitesiz ancak gürültünün şiddeti yüksek oranda bastırılarak elde edilmiştir.



Şekil 27 – İkinci test verisine üçüncü filtrenin uygulanması sonucunda çıkışta görülen genlik cevabı.

Şekil 28'de üçüncü test verisine 3 ayrı eğitim verisinde kullanılan filtreler uygulanmıştır. Birinci ve ikinci eğitim verisinde kullanılan filtrelerin gürültüyü yeteri kadar engelleyememesinden dolayı bu test verisinde de yine üçüncü eğitim verisinde kullanılan alçak geçiren filtre kullanılmıştır. Filtre çıkışında elde edilen seste gürültü olabildiğince azaltılsa da kalitesinde bozulmalar meydana gelmiştir.



Şekil 28 – Üçüncü test verisine üçüncü filtrenin uygulanması sonucunda çıkışta görülen genlik cevabı.

2.2.6. Sonuçların Analizi

Eğitim verilerindeki seslerin spektrum analizi sonucunda uygun filtreler tasarlanmış ve uygulanmıştır. Dolayısıyla eğitim verilerine uygulan filtreler sonucunda elde edilen seslerde yüksek oranda başarı sağlanmıştır. Fakat test verilerine eğitim verilerinde kullanılan filtreler uygulandığı için başarı oranı düşmüştür. Bunun sebebi her konuşmadaki insan seslerinin aynı frekanslar değerlerinde olmamasıdır. Bunda dolayı uygulanan filtreler test verilerinde istenmeye bozulmalara sebebiyet vermiştir. Bir diğer sebep ise sesin şiddeti ile gürültünün şiddetleri

arasındaki orandır. Gürültünün sese oranı artıkça filtreleme işlemi sonucunda elde edilen sesteki kalite azalır. Bu durumu engellemek için Wiener filtresi^[2] ve adaptif filtreleme^[3] yöntemlerinin kullanılabileceği görülmüş ancak ders kapsamından dışarıda olduğundan bu yöntemlerin uygulamasına girilmemiştir.

3. Sonuç ve Yorum

Proje kapsamında, Python kullanılarak ses verilerinin işlenmesi ve analiz edilmesi süreci üzerinde yoğunlaşmıştır. Bu süreçte spektrum analizi yöntemleri uygulanarak ses verilerinin ω -ekseninde bileşenleri incelenmiştir. Bu analiz sonuçlarına dayanarak etkili bir şekilde filtre tasarımı gerçekleştirilmiştir. Bu deneyim, ses işleme ve işaret işleme alanlarında ders kapsamındaki lisans düzeyinde bir anlayış geliştirmeye yardımcı olmuştur. Tasarlanan bu filtreler test verilerinde de uygulanmış ve belirli bir düzeyde başarı oranı sağladığı söylenebilir. Filtreleme yöntemi belirlenirken gürültünün tam olarak azaltılmayacağı tahmin edildiğinden sonuçlarda çok büyük sürprizlerle karşılaşılmasıdır. Önceki bölümlerde bahsedildiği üzere daha iyi sonuçlar elde etmek için Wiener filtresi^[2] ve adaptif filtreleme yöntemi^[3] düşünülmüştür ancak ders kapsamı dışında olduğundan bu yöntemler uygulanmamıştır. Ayrıca daha iyi sonuç elde etmek amacıyla seyrek örnekleme yapılarak frekans ekseninde genişleme amaçlanmıştır. Bu spektruma bakarak kesim frekansını daha iyi belirleyerek filtre tasarlamak istenmiştir ancak bu filtre uygulandığında beklenen sonuç elde edilmemiştir. Daha sonra çıkıştaki işaretlere sık örnekleme uygulayarak ses kalitesi artırılmak istenmiş fakat başarılı olunamamıştır. Farklı bir yöntem olarak da amplitude thresholding yöntemi^[4] düşünülmüş fakat konuşma ve gürültünün genlik değerlerinin birbirlerine yakın olmasından dolayı bu yöntem sağlıklı bir şekilde uygulanamamıştır. Yapay zeka, gürültülü seslerden gürültüyü kaldırmak için genellikle derin öğrenme modellerini (Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), U-Net) kullanarak, gürültülü ve temiz ses verileriyle eğitilir ve bu modeller yeni gürültülü sesleri işleyerek temizlenmiş ses sinyalleri üretir.

Kaynaklar

[1] en.wikipedia.org/wiki/Voice_frequency

[2] uol.de/f/6/dept/mediphsik/ag/sigproc/download/paper/s/doclo/journal_wiener.pdf

[3] B. Widrow, et al., “Adaptive Noise Cancelling: Principles and Applications”, Proc.IEEE, vol. 63, pp.1692-1716, Dec. 1975.

[4] lutpub.lut.fi/bitstream/handle/10024/164024/kandityo_Pauli_Anttonen.pdf?sequence=3&isAllowed=y