

ELM368 – DÖNEM PROJESİ

KEDİ VE AĞLAYAN BEBEK SESLERİNİN ANALİZİ

Alper Kaan İLİK, Mahmut Safa BULAT, Yağmur DERYA

151024037, 141024051, 171024011

akilik@gtu.edu.tr, safa.bulat2014@gtu.edu.tr, yagmur.derya2017@gtu.edu.tr

ABSTRACT (ÖZET)

Bu projedeki amaç kedi miyavlama sesi ile bebek ağlama sesinin analiz edilip birbirinden ayırt edilmesidir. Bu analizi yapmak için, sesler frekans domaininde incelenmiş bu inceleme sonucunda, $\omega = 0.25\pi$ frekansından sonra işaretlerin genlikleri arasında belirgin farklılıklar olduğu gözlemlenmiştir.

Gözlemlenen farklılıklardan yola çıkarak, kesim frekansı $\omega_c = 0.25\pi$ olan yüksek geçiren filtre tasarlanıp, ses işaretleri filtrelendikten sonra, enerjileri karşılaştırılmıştır. Karşılaştırma sonucu elde edilen (kedi ve bebek ağlama ses işaretlerinin) enerji verileri arasında 10 kat civarı fark olduğu gözlemlenmiştir. Bu sayede bu sesler belirli bir eşik değeri ile birbirlerinden kolayca ayrıştırılabilir hale gelmiştir.

ANAHTAR KELİMELELER

IIR Filtre, Fourier Dönüşümü, Spektrum, Enerji, Kedi Miyavlaması, Bebek Ağlaması

1. Giriş

Günümüzde, bebekli evlerin birçoğunda, ebeveynler ağlama sesine duyarlı birtakım cihazlar kullanmayı tercih etmektedir. Fakat, kedi miyavlama sesi sıklıkla bebek ağlama sesine benzetilmektedir. Bu yüzden, bebekli bir evde kedi de varsa, bu tarz cihazlar kedi sesi ile bebek ağlama sesini bir şekilde analiz edebilecek ve birbirinden ayırabilecek şekilde geliştirilmelidir.

Bu seslerin (kedi miyavlaması ile bebek ağlaması) frekans uzayında incelendiğinde de seslerin tizlik/baslık seviyeleri ve güçleri düşünüldüğünde çoğunlukla birbirlerine yakın genliklerde ve frekanslarda salındığı tahmin edilmektedir.

Bu projede amaç, yukarıda değinilen türde problemleri çözebilmek için, sesleri birbirinden ayırabilecek bir analiz yöntemi gerçekleştirmektir. Bu noktada, önceden de bahsedildiği gibi işaretlerin herhangi bir filtreleme işlemine uğramadan, sadece frekans uzayında incelenmesinin belki iki ses işareti için işe yarasa bile, genel bir analiz için yeterli olmayacağı düşünüldü.

Ses analizini frekans uzayında incelemenin yanı sıra, enerjilerini karşılaştırarak yapmak da mümkündür.

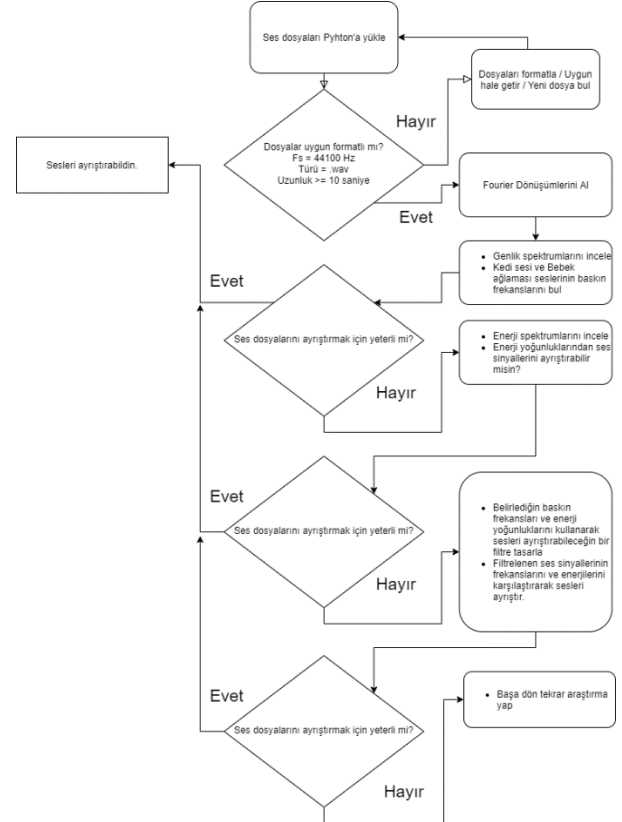
$$E = \sum_{n=-\infty}^{\infty} |x[n]|^2$$

Eşitlik 1

Enerji, eşitlik 1 ile hesaplanmakta olup, bu eşitlikten görüldüğü üzere, işaretin genliği ile ilişkilidir. Bu eşitlikten yola çıkarak, işaretlerin, enerji anlamında da birbirlerine yakın düzeylerde çıkması beklendiği için, frekans uzayındaki incelemelere benzer şekilde, işaretlerin filtrelenmeden sadece enerjilerine bakılarak ayırt edilemeyeceği görülmektedir.

Bu ön incelemeler sonucunda işaretlerin önce frekans uzayında, enerji yoğunlukları arasında belirgin farklılıkların olduğu bölge göz önünde bulundurularak, uygun bir filtre ile filtrelenmesi gerektiği düşünüldü. Daha sonra elde edilen (filtrelenen) işaretlerin enerji yoğunlukları karşılaştırıldığında, enerjileri arasında artık belirgin bir fark görülmesi beklenmektedir.

Elde edilen (kedi miyavlaması için üç, bebek ağlaması için üç olmak üzere toplam altı adet ve minimum 10 saniyelik) ses işaretleri kullanılarak, yapılan bu ön incelemelerin doğruluğu “Deneyler ve Analiz” bölümünde incelenecek ve problemin çözümü için gerekli analiz yöntemi gerçekleştirilecektir.



Şekil 1. Akış diyagramı.

2. Deneyler ve Analiz

Ön incelemeler ve araştırmalar sonucunda yapılmasına karar verilip, giriş bölümünde de bahsedilen analiz yönteminin akış diyagramı Şekil 1’de verildi.

2.1. Ses Verilerinin Elde Edilmesi

Kullanılması düşünülen 6 adet ses dosyası, günümüzün büyük problemi olan pandemi sebebi ile dışarı çıkıp kaydedilemeyeceği için, çevrimiçi platformlar yardımı ile elde edildi. Başlangıçta tercih edilen internet sitesindeki [1] ses dosyaları, deney için fazla gürültülü ve uygun formatta olmadığı için, günümüzün popüler yayın platformundan [2] sesler seçildi.

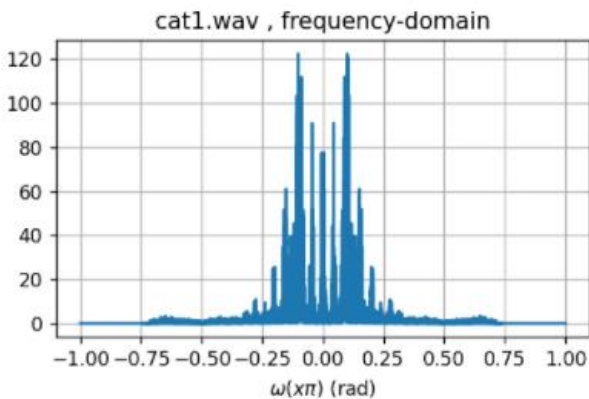
Elde edilen ses dosyaları, Python’da okuma ve analiz etme işlemlerinde çeşitli hatalar (dosya tipinin uygun olmaması, byte aşımı vb.) verdi. Bu hataların önüne geçilebilmesi ve verilerin Python’da rahatça kullanılabilmesi için, ses dosyaları, uzantısı “.wav” olan dosya türünde ve örnekleme frekansı $F_s = 44.1$ kHz olacak şekilde tekrar formatlandı.

2.2. Verilerin İlgili Ortamda İncelenmesi

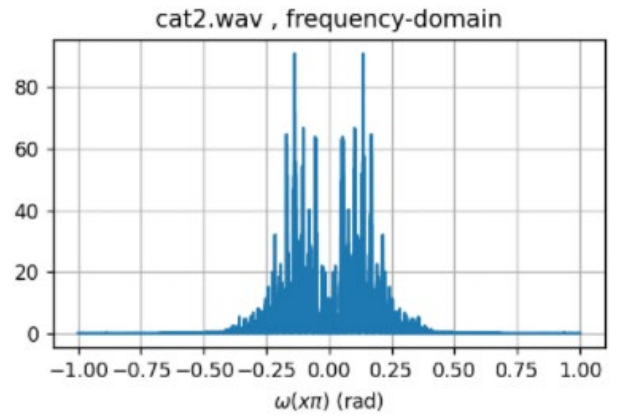
Formatlanan ses dosyaları “jupyter notebook”ta tanımlanarak, önce kedi miyavlama sesleri zaman uzayında ve Fourier dönüşümü kullanılarak frekans uzayında (Şekil 2.a, Şekil 2.b ve Şekil 2.c) daha sonra da bebek ağlama sesleri zaman uzayında ve frekans uzayında (Şekil 3.a, Şekil 3.b ve Şekil 3.c) çizdirildi.

Bahsedilen şekillerdeki grafikler incelendiğinde, işaretlerin spektrumlarının ufak farklılıklarla, genel anlamda benzer olduğu görüldü.

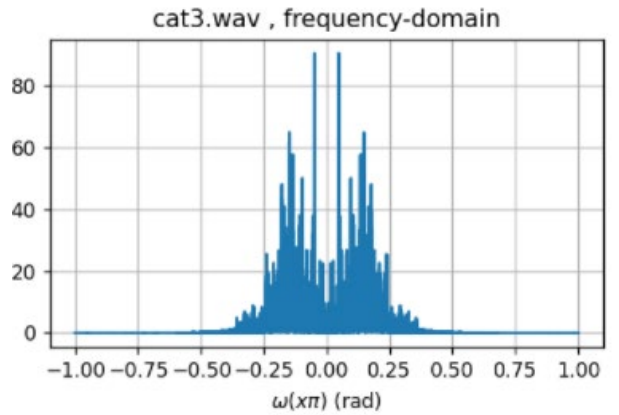
Kedi miyavlama seslerine ait olan Şekil 2.a-b-c’deki grafiklere bakıldığında, baskın işaretlerin ortak olarak alçak frekanslarda toplandığı, $\omega = \pm 0.25\pi$ ’den itibaren genliklerinin genele oranla oldukça düşük olduğu görüldü.



Şekil 2.a. Frekans uzayında ilk kedi miyavlama sesi.

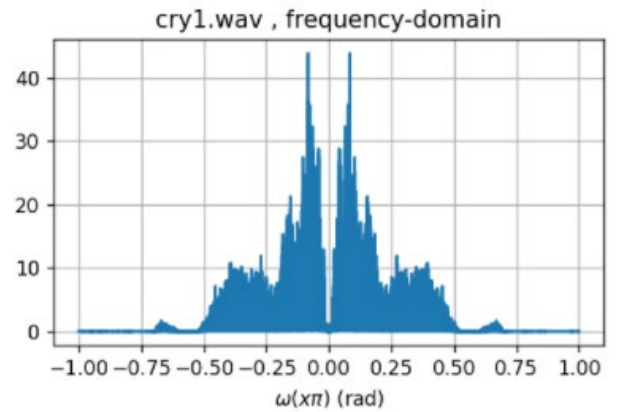


Şekil 2.b. Frekans uzayında ikinci kedi miyavlama sesi.

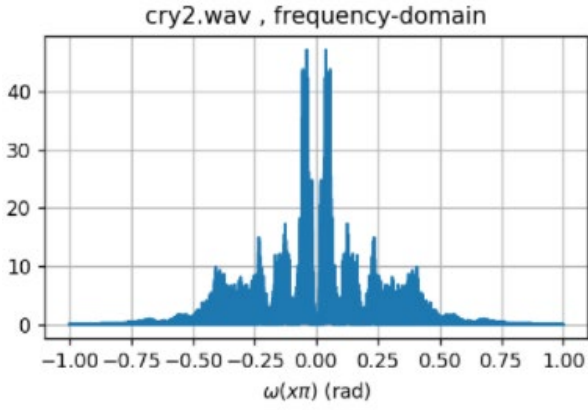


Şekil 2.c. Frekans uzayında üçüncü kedi miyavlama sesi.

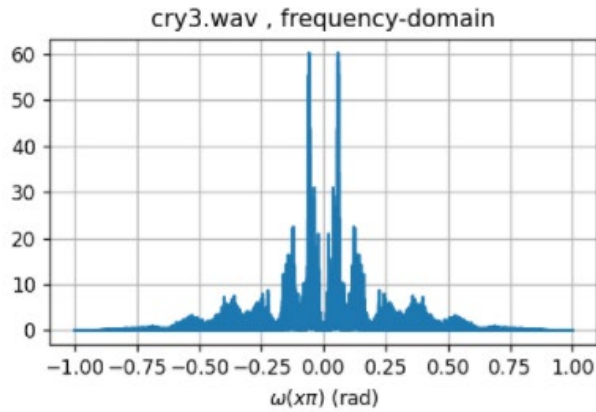
Bebek ağlama seslerine ait olan Şekil 3.a-b-c’deki grafiklerde ise yine baskın işaretlerin alçak frekanslarda olduğu fakat $\omega = \pm 0.25\pi$ ’den itibaren kedi miyavlama sesine kıyasla işaretlerin genliklerinin çok daha belirgin olduğu görüldü.



Şekil 3.a. Frekans uzayında ilk bebek ağlama sesi.



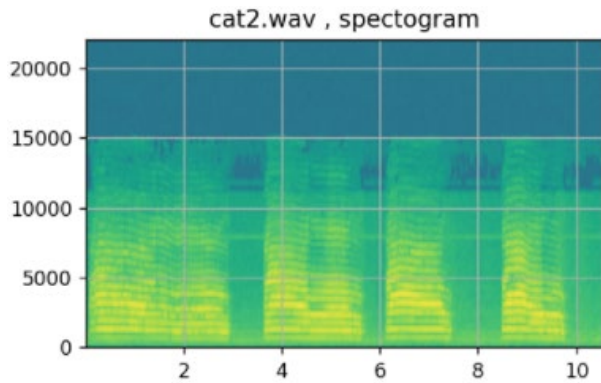
Şekil 3.b. Frekans uzayında ikinci bebek ağlama sesi.



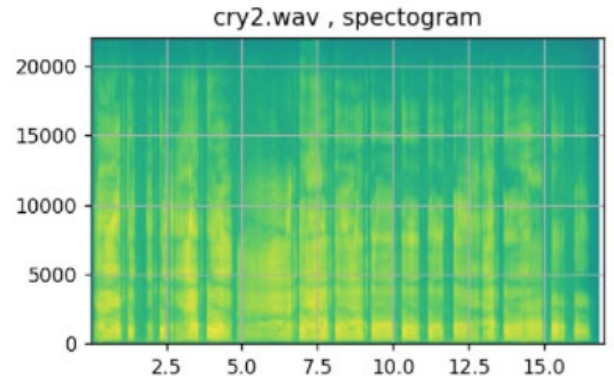
Şekil 3.c. Frekans uzayında üçüncü bebek ağlama sesi.

Elde edilen veriler, ön incelemelerde düşünüldüğü gibi olup, bir filtreleme işlemi yapmadan bu seslerin sadece frekans uzayındaki genliklerine bakılarak ayırt edilemediği görüldü.

Spektrogram; belirli bir dalga formunda bulunan çeşitli frekanslarda bir sinyalin gücünü veya yüksekliğini temsil eden görseldir [3]. Filtreleme işlemi yapmadan işaretlerin spektrogramları çizdirilerek zamanda frekanslarının nasıl değiştiği (Şekil 4.a-b) incelendi.



Şekil 4.a. İkinci kedi miyavlama sesi spektrogramı.



Şekil 4.b. İkinci bebek ağlama sesi spektrogramı.

Bu spektrogramlardan, bebek ağlama sesinin 20 kHz frekansının üzerine kadar çıktığı görülürken, kedi miyavlama sesinin maksimum 15 kHz frekansına kadar çıktığı görüldü. Fakat farklı ses dosyaları için inceleme yapıldığında bu durumun her zaman bu şekilde olmayıp, birbirleriyle neredeyse aynı spektrogramlara sahip olabilecekleri de gözlemlendi. Bu sebeple filtreleme işlemi yapmadan bu seslerin analizinin her zaman doğru şekilde yapılamayacağı düşünüldü. Yine de emin olmak adına işaretlerin enerjileri hesaplandı.

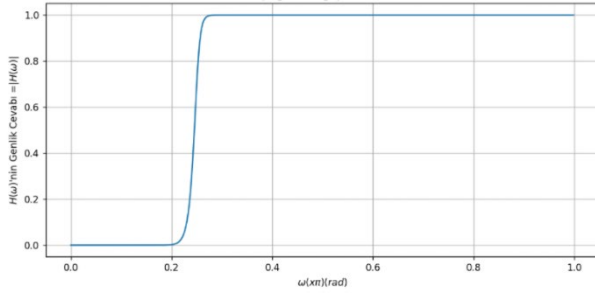
Kedi Miyavlama Sesi	8.16×10^{18}	2.53×10^{18}	2.62×10^{18}
Bebek Ağlama Sesi	2.85×10^{18}	3.49×10^{18}	2.93×10^{18}

Tablo 1. Ses işaretlerinin enerji (joule) değerleri.

Tablo 1’de görüldüğü üzere, ses dosyalarının enerji değerleri filtreleme yapılmadığı durumda birinci kedi miyavlama sesi hariç birbirlerine oldukça yakın çıktı. Bu durumda belirgin bir ayrımın yapılamayacağı açıkça görülmektedir.

2.3. Filtrenin Tasarlanması

Frekans uzayında yapılan incelemeler sonucunda $\omega = \pm 0.25\pi$ değerinden sonra işaretlerin genliklerinin oranlarında belirgin farklılıklar olduğu görülmüştü. Genlik ve enerji birbirleriyle ilişkili oldukları için, işaretleri, belirgin farklılıkların olduğu bölgelere göre filtreleyecek şekilde bir filtre tasarlanıp, bu filtreden geçirilirse enerji düzeyinde belirgin farklılıklar elde edilebileceği görüldü. Baskın işaretler düşük frekanslarda toplandığı için yüksek geçiren filtre tasarlanarak daha keskin sonuçlar elde edilebileceği düşünüldü.

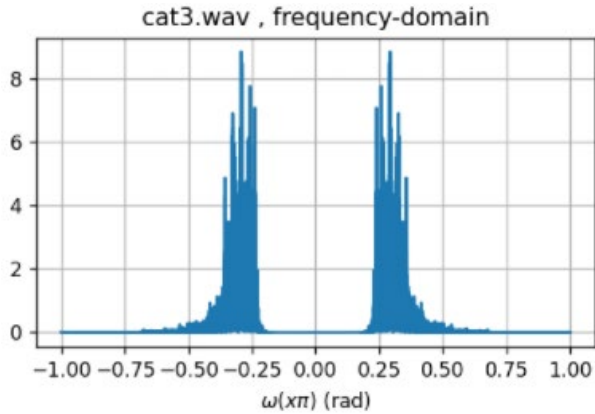


Şekil 5. Tasarlanan yüksek geçiren filtre.

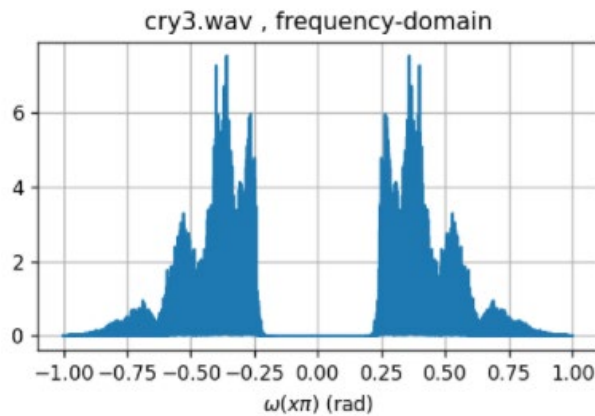
Şekil 5’te görülen, kesim frekansı $\omega = 0.25\pi$ olan IIR Butterworth filtresi Pyfda üzerinde tasarlandı. Tasarım yapılırken, filtrenin ideale yakın olması için, filtre parametreleri; $N = 25$, $F_{SB} = 0.119$, $F_{PB} = 0.125$ olarak seçildi. Daha sonra, tasarlanan filtre “scipy” kütüphanesindeki “loadmat” komutu ile Python’da açıldı. Tasarlanan filtreden geçtikten sonra işaretlerin enerjileri arasındaki farkın daha belirgin olması beklenmektedir.

2.4. Filtrelenen İşaretlerin İncelenmesi

Filtreden geçtikten sonra, $\omega = (0, 0.25\pi)$ arasındaki işaretlerin filtrelenmesi beklenmekteydi. Şekil 6.a-b.’de görüldüğü üzere, filtreleme işlemi beklendiği gibi başarı ile gerçekleştirildi.



Şekil 6.a. Filtrelenmiş kedi miyavlama sesi.



Şekil 6.b. Filtrelenmiş bebek ağlama sesi.

İşaretler filtrelendikten sonra kedi miyavlama sesleri ile bebek ağlama seslerinin mutlak genlik toplamlarının belirgin bir şekilde değiştiği, Şekil 6.a ve Şekil 6.b

karşılaştırıldığında, kedi miyavlama sesinin, bebek ağlama sesine göre genlik bileşenlerinin çok daha az olduğu görülmektedir.

Bu aşamada, filtrelenen işaretlerin genliklerine bakarak işaretleri ayırtmak mümkündür fakat bu genlik değerlerinin karelerini inceleyerek daha genel ve belirgin bir fark elde edilmesi mümkündür. İşaretlerin enerjilerinin genlik toplamının karesine eşit olduğu (bkz. eşitlik 1) bilindiğine göre bu aşamada işaretlerin enerjilerinin hesabı yapılarak bu işaretlerin birbirlerinden daha net bir şekilde ayrılabilceği düşünüldü.

2.5. Filtrelenen İşaretlerin Enerjilerinin İncelenmesi

İşaretler filtrelendikten sonra, yukarıda teorik olarak düşünülen enerji ile işaretlerin ayrıştırılması için, filtrelenen işaretlerin Fourier dönüşümü alınarak, elde edilen bu işaretlerin enerjileri, eşitlik 1 kullanılarak, frekans uzayındaki genlikleri ile hesaplandı.

Kedi Miyavlama Sesi	6.40×10^{16}	3.88×10^{16}	5.05×10^{16}
Bebek Ağlama Sesi	4.72×10^{17}	5.23×10^{17}	2.96×10^{17}

Tablo 2. Filtrelenen ses işaretlerinin enerji (joule) değerleri.

Tablo 2’ye bakıldığında, daha önceden de düşünüldüğü gibi filtreleme işleminden geçtikten sonra işaretlerin enerjileri arasında belirgin bir fark oluşmuş olup bu fark 10 kat civarındadır. Enerjileri arasındaki bu fark, işaretleri birbirinden ayırtmak için yeterli büyüklüktedir ve 1×10^{17} civarında bir eşik değeri seçildiğinde enerjisi eşik değerinin altında olan işaretin kedi miyavlama sesine, üstünde olan işaretin ise bebek ağlama sesine ait olduğu yorumu yapılabilir. Tasarlanan filtre ile işaretlerin filtrelenmesinden sonra seslerin ayrımı yapılabildi.

3. Sonuç ve Yorum

Kedi miyavlama sesi ile bebek ağlama sesini birbirinden ayıracak analiz yöntemi başarıyla gerçekleştirildi. Bu yöntem geliştirilirken, işaretlerin frekans uzayında incelenmesinin yeterli olup olmayacağı üzerine net düşüncelere sahip değildik. İşaretlerin birbirlerine benzerliği sebebiyle frekans uzayında benzerliklerin olacağı beklenmekteydi fakat frekansların birbirlerine bu kadar yakın olacağı düşünülmemişti. Frekans uzayındaki bu benzerlikler enerji analizinde de oldukça yakın sonuçlar elde edilmesine sebep oldu. Elde edilen sonuçlar doğrultusunda, ses işaretlerinin herhangi bir işleme uğramadan sadece frekans uzayında incelenmesinin yetersiz olduğu görüldü.

Bu gözlemler sonucunda anlamlı farklılıkların olduğu aralığın belirlenip, işaretlerin filtrelenerek bu analizin yapılabileceği görüldü.

Kullanılan filtre ideale yakın bir IIR filtre olduđu için maliyet bir miktar artacaktır. Bu çözümün geçerliliđi için, ses işaretlerinin yüksek kalitede (gürültüsü az) ve uygun formatta olması gerekmektedir.

Çözümü geliştirmek için, giriş işaretlerini, maksimum genlik değerlerini belirli bir değere getiren bir filtreden geçirerek başlangıç şartları belli bir standartta tutulabilir. Bu şekilde enerji hesabı yapılırken, genliđi çok yüksek olan bir sesin enerjisinin de çok yüksek olmasının önüne geçilirken, aynı zamanda işaretlerin frekans uzayındaki dağılımları oransal olarak değışmemiş olacaktır.

Kaynaklar

[1]<https://freesound.org>

[2]<https://www.youtube.com>

[3]<https://medium.com/datarunner/librosa-9729c09ecf7a>

[4]<https://pythontic.com/visualization/signals/spectrogram>

[5]<https://stackoverflow.com/questions/2060628/reading-wav-files-in-python>