

## 1. Problem Tanımı

Günlük hayatta kaydedilen sesler, çevresel gürültülerle kirlenebilmektedir. Bu projede amaç, farklı türdeki arka plan gürültülerini otomatik olarak tespit edip bastırarak temiz ve anlaşılır ses elde etmektir.

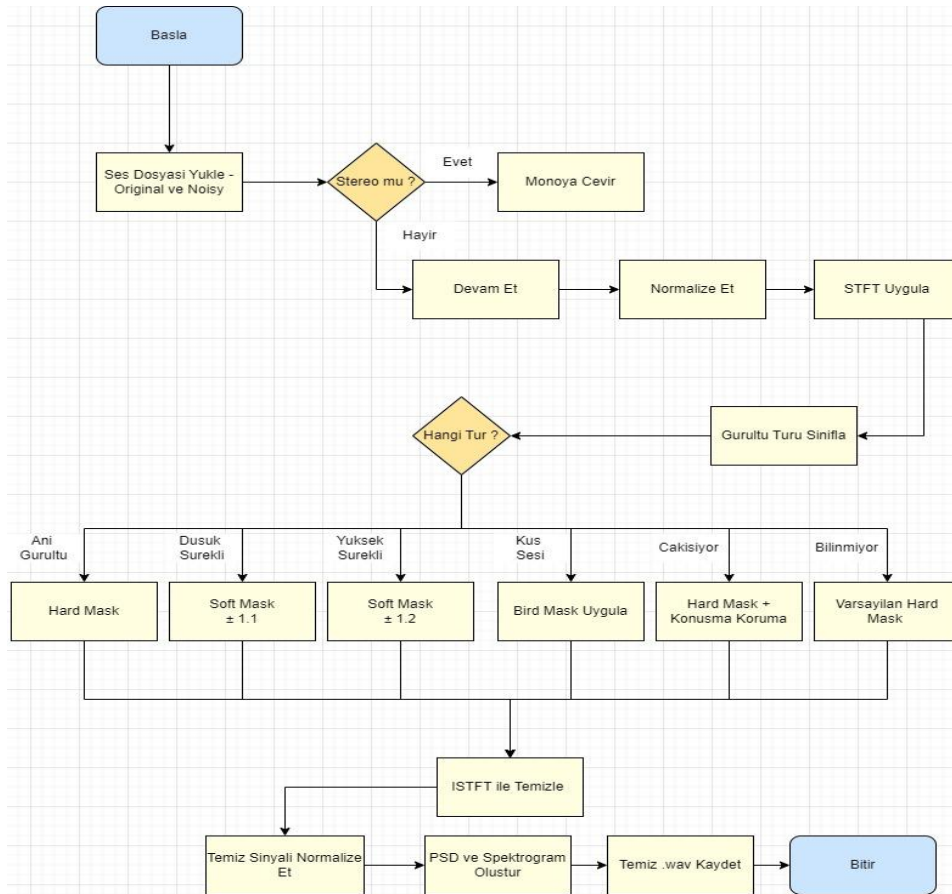
## 2- Algoritma Tasarımı

### 2.1. Kullanılan Kütüphaneler:

Kütüphane	Açıklama
NumPy	Sayısal işlemler, matris hesaplamaları ve diziler için kullanıldı.
SciPy (scipy.io, scipy.signal)	Ses dosyalarını okuma/yazma (wavfile), STFT ve iSTFT dönüşümleri için kullanıldı.
Matplotlib	PSD ve spektrogram grafiklerini çizmek ve analiz etmek için kullanıldı.
os	Dosya işlemleri ve klasör oluşturma gibi sistemsel görevler için kullanıldı.

Tablo 1-kullanılan kütüphaneler

### 2.2. Akış Diyagramı:



Şekil 1-Akış Diyagramı (Draw.io ile el ile çizerek hazırlanmıştır.)

### 2.3. Başlangıç ve Sinyal Hazırlığı

Projemizdeki ilk adım, elimizdeki iki ses dosyasını (biri temiz, biri gürültülü) işleme almaktır. Eğer sesler stereo ise, analiz sürecini sadeleştirmek için mono'ya çevrilir. Bu sayede filtreleme doğruluğunu artırırız ve kanallar arası dengesizlik kaynaklı hatalardan kaçınırız. Ardından her iki sinyal normalize edilir. Bu, ses seviyelerini dengeler ve filtreleme sırasında bozulmaları önler. Böylece orijinal ve gürültülü sinyalleri eşit koşullarda karşılaştırmak mümkün olur. Normalleştirme şu şekilde yapılır:

$$x_{\text{norm}}[n] = \frac{x[n]}{\max(|x[n]|) + \varepsilon}$$

Bu sayede sistem, analiz ve filtreleme için sağlıklı bir başlangıç yapmış olur. Tüm bu işlemler tamamlandığında elimizde artık hem “orijinal”, hem “gürültülü”, hem de **zaman-frekans haritası çıkarılmış** sinyaller bulunur.

#### 2.4. Frekans Analizi (STFT) ve Gürültü Tanıma

Tiz bir kuş sesiyle, derin bir motor uğultusu zaman ekseninde benzer görünebilir; ama frekans ekseninde çok farklıdır. Bu yüzden sinyali hem zamana hem frekansa göre analiz eden **STFT (Short-Time Fourier Transform)** yöntemini kullandık. Kısaca, STFT, sesi küçük zaman dilimlerine böler ve her bölümde hangi frekansların baskın olduğunu gösterir.

$$X(n, \omega) = \sum x[m] \cdot w[n - m] \cdot e^{-j\omega m}$$

Burada  $w[n]$  kısa bir pencere fonksiyonudur. Bu adımda, gürültülü sesin **enerji** ve **frekans dağılımı** incelenerek gürültü tipi otomatik olarak belirlenir. Öne çıkan ölçümler arasında; **spektral ağırlık merkezi**, frekansların ortalama yoğunluğunu temsil ederken; **enerji dalgalanması (standart sapma)**, sinyalin zaman içindeki değişkenliğini göstermekte; **baskın frekans** ise sinyalde en yüksek enerjiye sahip frekans bileşenini ifade etmektedir.

Tür	Özellik
Ani Gürültü	Ani ve yüksek frekans artışı
Çarpışan Sesler	Dalgali enerji + orta frekans
Kuş Sesi / İnce Tiz	Çok yüksek frekans (>6000 Hz)
Sürekli Düşük	Sabit ve düşük frekans (<600 Hz)
Sürekli Orta	600–1500 Hz arası sabit enerji

#### 2.5. Gürültüye Uygun Maske Uygulama

Gürültü Türü	Maske Türü	Açıklama
Ani (havlama, çarpma)	Hard Mask	Eşik değerinin altında kalan frekanslar tamamen bastırılır.
Sürekli Düşük	Soft Mask (düşük eşik)	Gürültüye karşı daha yumuşak bastırma uygulanır, ses doğallığı korunur.
Sürekli Orta	Soft Mask (orta eşik)	Biraz daha güçlü bastırma yapılır.
Kuş sesi	Bird Mask	2000–8000 Hz arası frekanslar özel olarak yumuşatılır.
Çarpışma + Konuşma	Hard + Konuşma Koruma	Konuşma frekansları (300–3400 Hz) korunur.

Maskeleme Formülasyonu (Soft örneği):

$$\text{gain}(f) = \min \left( 1, \frac{|X_{\text{clean}}(f)|}{|X_{\text{noisy}}(f)| \cdot \text{eşik}} \right)$$

Bu katsayı frekans uygulanarak zayıf sinyaller bastırılır, güçlü sinyaller korunur. Bu adımda önemli olan, konuşma içeriğinin bozulmaması için agresif filtreleme yerine akıllı maskeleme uygulanmasıdır. Filtreleme sonrasında ters STFT (iSTFT) ile sinyal tekrar zaman düzlemine çevrilir.

#### 2.6. Ters Dönüşüm ve Çıkış Üretimi

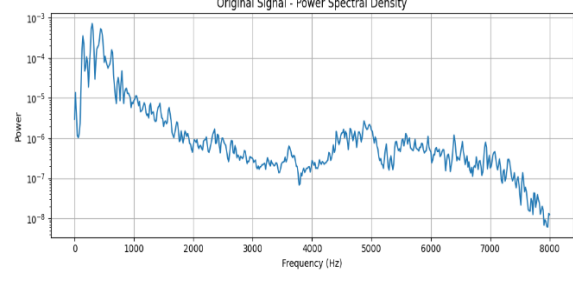
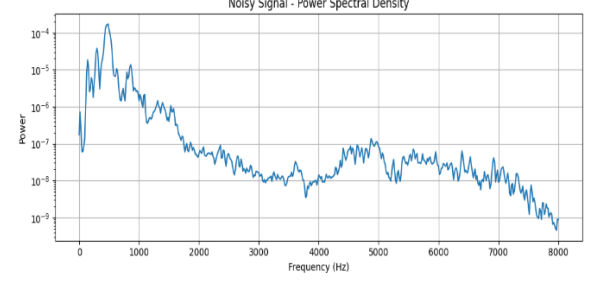
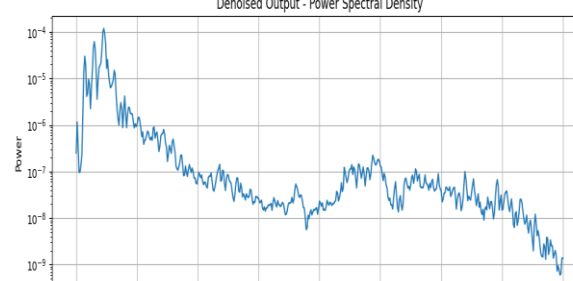
Filtrelenmiş (maskelenmiş) sinyalin frekans-temsilinden tekrar zaman düzlemine dönüştürülmesi için **ters kısa süreli Fourier dönüşümü (iSTFT)** uygulanır.

Bu dönüşüm sayesinde, daha önce gürültü azaltma işlemlerine tabi tutulmuş sinyal, dinlenebilir ve kaydedilebilir hale gelir. Ardından sinyal normalize edilerek .wav formatında dışa aktarılır. Böylece, sinyal seviyesi dengelenir (bozulmaları önler), dinleyici için kullanıma hazır bir sonuç elde edilir ve grafiksel analiz (PSD ve spektrogram) çıktıları da üretilerek kalite kontrolü yapılır.

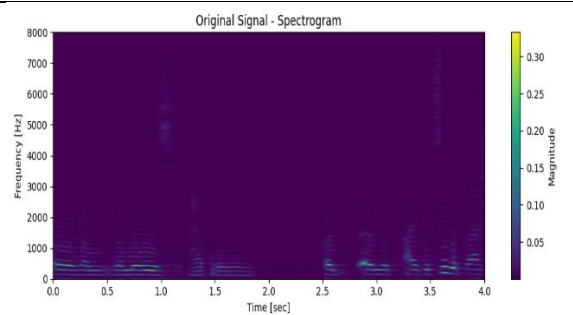
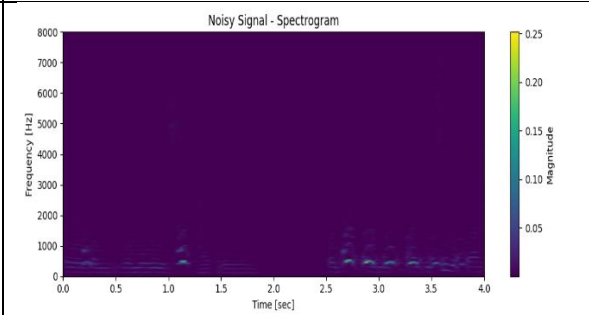
### 3. Sinyal Analizi ve Nedensel Karşılaştırma

Bu kısımda, gürültülü ve filtrelenmiş (temizlenmiş) sinyallerin Power Spectral Density (PSD) ve spektrogram çıktıları üzerinden karşılaştırmalı bir analizi sunulmuştur.

#### 3.1 Güç Spektral Yoğunluğu (PSD) Analizi

Orijinal Sinyal– PSD	Gürültülü Sinyal – PSD
	
Bu görsel referans amaçlıdır. Enerji dağılımı dengelidir ve konuşma bandı (300–3400 Hz) net olarak gözlemlenmektedir.	Grafikte özellikle 0–800 Hz aralığında anormal düzeyde yüksek bir enerji görülmektedir. Bu, sinyale düşük frekanslı sürekli bir gürültünün eklendiğini göstermektedir. Bu tür bir gürültü, konuşma bandına (300–3400 Hz) giren düşük frekans bileşenleri maskeleyerek anlaşılabilirliği düşürür. PSD'deki bu tepe, sinyalin baskılandığını ve dengesiz hale geldiğini doğrudan göstermektedir.
Temizlenmiş Sinyal – PSD	
	Filtreleme sonrası 0-800 Hz arasında ciddi bir enerji azalışı gözlemlenmektedir. Düşük frekanslardaki aşırı yük ortadan kaldırılmış; böylece sinyal, konuşma frekanslarında daha dengeli hale gelmiştir. Bu değişim, uygulanan "soft_mask" filtresinin hedeflenen frekans aralığında başarılı bir şekilde çalıştığını göstermektedir. Ayrıca bu filtre, konuşma içeriğini koruyarak sadece gürültüyü bastırdığı için ideal bir çözüm sağlamıştır.

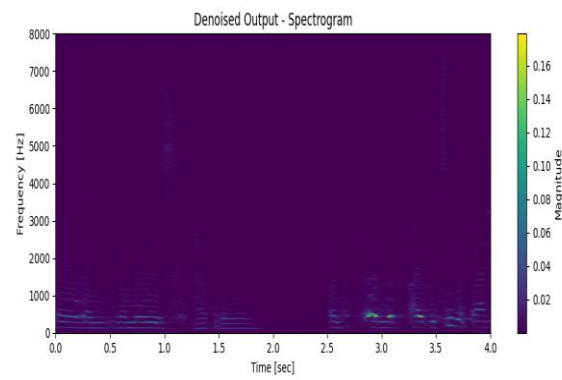
#### 3.2 Spektrogram Analizi

Orijinal Sinyal– Spektrogram	Gürültülü Sinyal – Spektrogram
	
Bu spektrogramda, konuşmaya özgü ses bileşenleri olan <b>harmonik çizgiler</b> açık şekilde seçilebilmektedir. Harmonikler, temel bir	Spektrogramda, zaman boyunca sabit kalan ve <b>0–1000 Hz</b> arasında yoğunlaşan enerji yığılması dikkat çekmektedir. Bu sürekli yapı, gürültünün

frekansın katları olarak oluşan ve özellikle insan sesinde belirgin olan frekans yapılarıdır. Bu yapıların net olması, sinyalin temiz kaydedildiğini ve konuşma bilgilerini içerdiğini gösterir. Görseldeki parlak bölgeler ise zaman içinde enerjinin yoğunlaştığı frekansları temsil eder; bu da sesin güçlü olduğu anları yansıtır.

yalnızca varlığını değil, konuşma detaylarını nasıl bastırdığını da açıkça göstermektedir. Harmonik yapıların silikleşmesi, konuşma sinyalinin gürültü altında ezildiğini göstermektedir.

#### Temizlenmiş Sinyal – Spektrogram



Temizleme işlemi sonrasında aynı frekans aralığındaki yoğunluk büyük oranda azalmış; sinyale ait harmonikler yeniden görünür hale gelmiştir. Bu durum, yalnızca gürültünün değil, onun etkilediği bölgelerin de başarıyla geri kazanıldığını ortaya koymaktadır. Zaman–frekans düzeyinde netlik artmış, konuşma bilgisi daha ayırt edilebilir bir hale gelmiştir.

#### 4.Sonuç ve Değerlendirme

Tasarladığımız sistem, konuşma sinyalini bozmadan arka plan gürültülerini iyi bir doğrulukla bastırmıştır. STFT ile ses zamana ve frekansa göre analiz edilmiş, belirgin frekanslardaki gürültüler özel maskelemelerle temizlenmiştir. Grafikler, sesin yapısını görsel olarak değerlendirmeyi sağlamış; havlama, kuş sesi, siren ve motor gibi yaygın gürültüler etkili biçimde azaltılmıştır. Ancak özellikle **hibrit gürültülerin ve sınıflandırılmayan (unknown) ses türlerinin**, sistemin karar mekanizmasını zorladığını fark ettik.

- Aynı anda birden fazla türde gürültü içeren sesler **hibrit gürültüler** olarak adlandırılır. Örneğin bir ses dosyasında hem düşük frekanslı sabit bir uğultu (örneğin fan sesi) hem de yüksek frekanslı ani bir gürültü (örneğin kapı çarpması) yer alıyorsa, bu iki farklı yapıyı tek bir filtreyle etkili şekilde bastırmak zordur. Bu sorunu aşmak için, STFT analizi ile elde edilen frekans-temelli veriler üzerinden **bölgesel maskeleme** yaklaşımı uygulanabilir. Yani her frekans aralığı için farklı bir maskeleme stratejisi uygulanır: örneğin 0–600 Hz aralığında sürekli gürültüler için “soft mask”, 6000 Hz üzerindeki tiz ve ani sesler için ise “hard mask” kullanılır. Böylece hibrit ortamlarda konuşma sesi korunurken, farklı karakterdeki arka plan gürültüleri etkili biçimde bastırılabilir.
- Unknown olarak tanımlanan gürültüler, sistemin daha önce karşılaşmadığı veya belirgin bir frekans yapısına sahip olmayan sesleri ifade eder. Örneğin çocuk kahkahası, alkış gibi sesler klasik frekans kalıplarına uymadığından sistem tarafından doğru şekilde sınıflandırılmaz ve bu da ya gereğinden fazla bastırmaya ya da yetersiz filtrelemeye yol açabilir. Bu gibi durumlarda konuşmanın doğallığı zarar görebilir. Bu sorunu aşmak için, sistemin bu tür bilinmeyen seslere karşı daha esnek yaklaşması hedeflenmektedir. Bu durum için bu sesleri geçici olarak “unknown” sınıfına alıp düşük etkili bir **soft-pass filtresi** ile işlemeyi düşündük. Böylece sinyalin genel yapısı bozulmaz. Aynı zamanda bu bilinmeyen sesler bir veri kümesinde toplanarak, gelecekte **SVM** veya **KNN** gibi öğrenme tabanlı algoritmalarla yeniden sınıflandırılabilir. Bu iyileştirme sayesinde sistem, zamanla kendi deneyimlerinden öğrenerek farklı gürültü türlerini tanıma ve uygun maskeleme uygulama yeteneğini geliştirebilir.



**Hazırlayanlar:**

Zehra Demirtaş 23010903135

Alaa Lafi B210109602