

Araştırma Makalesi / Research Article**Ses Kaydı Yapılan Görüşmelerde Arka Plan Gürültüsü Azaltımı: Kayıt ve Kayıt Sonrası İşleme Stratejileri***Background Noise Reduction In Audio-Recorded Interviews: Recording and Post-Processing Strategies*Mustafa DEMİR¹ , Kadri Yılmaz ERDAL²  & İnan GÜMÜŞ³ DOI : [10.63556/ankad.v10i3.471](https://doi.org/10.63556/ankad.v10i3.471)

Geliş/Received: 10/03/2026

Kabul/Accepted: 18/05/2026

Öz

Bu çalışma, nitel araştırmalarda veri toplama sürecinde sıkça karşılaşılan ses kayıt kalitesi sorunlarına yönelik teknik ve metodolojik çözüm önerileri geliştirmeyi amaçlamaktadır. Özellikle saha ortamlarında gerçekleştirilen görüşmelerde kullanılan ses kayıtları; çevresel gürültü, mikrofon yerleşimi, akustik koşullar ve konuşmacıya özgü değişkenlerden etkilenmekte, bu durum verinin güvenilirliğini ve analiz edilebilirliğini azaltmaktadır. Çalışmada, kayıt öncesi ve sonrası süreçler için uygun mikrofon seçimi, temel akustik düzenlemeler ve filtreleme stratejilerine ek olarak Python ortamında geliştirilen dijital sinyal işleme (DSP) ve yapay zekâ (AI) temelli hibrit bir ses iyileştirme sistemi sunulmuştur. Bu doğrultuda sistem, enerji ve spektral düzlük temelli adaptif eşikleme, uğultu ve klik/plosive temizliği, tam-adaptif spektral gürültü azaltma ve konuşmayı ön plana çıkarma adımlarından oluşmakta ve ayrıca Demucs tabanlı kaynak ayırıştırma ve Whisper modeli ile gerçekleştirilen otomatik Türkçe transkripsiyon modülünü içermektedir. Sistem performansı, araştırmacı tarafından oluşturulan referans transkripsiyon ile model çıktısının karşılaştırılması yoluyla değerlendirilmiştir. Bu doğrulama sürecinde Kelime Hata Oranı (Word Error Rate, WER) ve Karakter Hata Oranı (Character Error Rate, CER) metrikleri hesaplanmış; Türkçe karakterlere yönelik normalizasyon işlemleri uygulanmıştır. Örnek doğrulama sonuçlarına göre sistem %79-%95 kelime doğruluğu ve %88-%97 karakter doğruluğu sağlamış; WER değeri 0.203 ve CER değeri 0.114 olarak hesaplanmıştır. Bulgular, geliştirilen hibrit yapının hem ses iyileştirme hem de otomatik transkripsiyon doğruluğu açısından nitel araştırmalarda kullanılabilir bir teknik çerçeve sunduğunu göstermektedir.

Anahtar Kelimeler: Görüşme tekniği, ses kaydı, arka plan gürültüsü, gürültü azaltma, dijital ses işleme.

Abstract

This study aims to develop technical and methodological solution recommendations for audio recording quality problems frequently encountered during the data collection process in qualitative research. Audio recordings used in interviews, especially in field settings, are affected by environmental noise, microphone placement, acoustic conditions, and speaker-specific variables, which reduce the reliability and analyzability of the data. In addition to appropriate microphone selection, basic acoustic arrangements, and filtering strategies for the pre- and post-recording processes, a hybrid audio enhancement system based on digital signal processing (DSP) and artificial

¹Sorumlu Yazar/Corresponding Author, Dr. Öğr. Üyesi, Burdur Mehmet Akif Ersoy Üniversitesi Türk Müziği Devlet Konservatuarı, Burdur, Türkiye. mdemir@mehmetakif.edu.tr, 0000-0001-8429-1113.

²Doç., Burdur Mehmet Akif Ersoy Üniversitesi Türk Müziği Devlet Konservatuarı, Burdur, Türkiye. kyerdal@mehmetakif.edu.tr, 0000-0003-1257-1751.

³Doç. Dr., Isparta Uygulamalı Bilimler Üniversitesi Uluborlu Selahhatin Karasoy Meslek Yüksekokulu, Isparta, Türkiye. inangumus@isparta.edu.tr, 0000-0002-6124-8881.

Önerilen Atıf/Suggestion Citation

Demir, M., Erdal, K.Y. & Gümüş, İ. (2026). Ses Kaydı Yapılan Görüşmelerde Arka Plan Gürültüsü Azaltımı: Kayıt ve Kayıt Sonrası İşleme Stratejileri. *Anadolu Kültürel Araştırmalar Dergisi*, 10(3), 1542-1558.

intelligence (AI) developed in a Python environment is presented. Accordingly, the system consists of steps including energy- and spectral flatness-based adaptive thresholding, hum and click/plosive cleanup, fully adaptive spectral noise reduction, and speech foregrounding. Also, it includes a Demucs-based source separation module and an automatic Turkish transcription module implemented with the Whisper model. System performance was evaluated by comparing the model output with a reference transcription created by the researcher. In this validation process, Word Error Rate (WER) and Character Error Rate (CER) metrics were calculated, and normalization procedures for Turkish characters were applied. According to the sample validation results, the system achieved 79%–95% word accuracy and 88%–97% character accuracy; WER was 0.203, and CER was 0.114. The findings show that the developed hybrid structure provides a usable technical framework for qualitative research in terms of both audio enhancement and automatic transcription accuracy.

Keywords: interview, audio recording, background noise, noise reduction, digital signal processing.

1. GİRİŞ

Paradigmalar, gerçekliği anlamanın, bilgi oluşturma ve dünya hakkında bilgi toplamanın tercih edilen yollarıdır. Bir araştırmacının paradigması; ontoloji (gerçekliğin doğası), epistemoloji (bilginin doğası), aksiyoloji (araştırma ve teorileştirme alanlarıyla ilişkili değerler) veya metodoloji (veri toplama, derleme ve analiz etme stratejileri) temelinde farklılık gösterebilir. İnsanlar bu konularda farklı duruşlar sergilediğinden paradigmaları oluşturan temel argümanları ve bakış açılarını anlamak önemlidir. Farklı disiplinler, paradigmalar için pozitivist/post-pozitivist, yorumlayıcı, eleştirel ve postmodern/postyapısallık gibi farklı terimler ve kategorilendirme şemaları kullanır (Tracy, 2013). Bu nedenle araştırma yaklaşımları farklılık gösterse de bilimsel araştırmalar, bilgi üretmeyi amaçlayan sistemli ve yapılandırılmış süreçleri ifade eder. Bilimsel yöntem; sorun, hipotezler (veya sorular), veri toplama, sonuçlar ve tartışmayı içeren bir süreç olarak tanımlanabilir. Süreç bir konu veya sorunla başlar, sorunla ilgili literatür incelenir, sorular sorulur, veriler toplanır ve analiz edilir, ardından raporlar yazılır (Creswell, 2013). Bu çalışma, nitel görüşme verilerinin teknik kayıt koşullarından kaynaklanan kalite sorunlarının, veri analizi ve transkripsiyon doğruluğu üzerindeki etkisini ele almaktadır. Özellikle saha kayıtlarında ortaya çıkan gürültü, akustik bozulma ve çevresel parazitlerin sistematik biçimde iyileştirilmesi, verinin metodolojik güvenilirliği açısından belirleyici bir değişken olarak değerlendirilmektedir. Bu doğrultuda, dijital sinyal işleme ve yapay zekâ destekli bir iyileştirme hattı önerilmektedir.

Bilimsel araştırmalarda bir olgu veya duruma ilişkin yapılmış gözlem, araştırmacı tarafından bir hipotez kurularak sınanmak istenir. Bu noktada araştırmacının çalışma disiplini doğrultusunda verilere ulaşabilmesi için, diğer bir deyişle hipotezin sınanabilmesi için bir araştırma modeli seçilir ve soruna ilişkin bir veri toplama yöntemi belirlenir. Araştırmanın geçerliliği ve güvenilirliği, büyük ölçüde elde edilen verilerin niteliğine ve bu verilerin toplandığı araçların uygunluğuna bağlıdır. Bu bağlamda araştırmacılar merak edilen konuya ilişkin bilgi elde edebilmek için çeşitli veri toplama tekniklerine başvurur. Araştırmacının hangi veri toplama tekniğini kullanacağı, araştırmanın problemine, verinin özelliğine veya veri kaynağına bağlı olarak değişebilmektedir. Alanyazında farklılıklar olsa da veri toplama yaklaşımları genel olarak yazılı veri toplama (anket, ölçek), görüşme, gözlem ve doküman incelemesi biçiminde sınıflandırılmaktadır (Büyüköztürk vd., 2018; Karasar, 2011; Yıldırım & Şimşek, 2018).

Nitel araştırma modellerine göre değişiklik gösteren veri toplama yöntemi ve uygulaması, olgu veya durumun en az hatayla ortaya konması bakımından önemlidir. Nitel metodoloji ifadesi en geniş anlamıyla, insanların kendi yazılı veya sözlü ifadeleri ve gözlemlenebilir davranışları olan betimsel veriler üreten araştırmaları ifade eder (Taylor, Bogdan & DeVault, 2016). Creswell (2013), bu verilerin değişkenliği durumunu şu şekilde açıklar: Veri toplama, alana uygunluğu açısından (örneğin, etnografide daha fazla gözlem, temellendirilmiş teoride daha fazla görüşme) ve veri toplama kapsamı açısından (örneğin, fenomenolojide sadece görüşmeler, vaka çalışması araştırmasında çoklu formlar vb.) farklılık gösterir. Veri analizi aşamasında farklılıklar en belirgin şekilde ortaya çıkar. Farklılık sadece analiz aşamasının özgüllüğüyle (örneğin, temellendirilmiş teori en özgül, anlatı araştırması daha az tanımlı) sınırlı değildir, aynı zamanda gerçekleştirilecek adımların sayısı da değişir (örneğin, fenomenolojide kapsamlı adımlar, etnografide az sayıda adım). Her yaklaşımın sonucu olan yazılı rapor, önceki tüm süreçlerden şekillenir. Bir bireyin hayatıyla ilgili hikâyeler, anlatı araştırmasını oluşturur. Genellikle görsel bir modelde tasvir edilen teori, temellendirilmiş teoride ortaya çıkar ve bir kültür

paylaşan grubun nasıl çalıştığına dair bütünsel bir bakış, etnografi ile sonuçlanır. Sınırlı bir sistemin veya bir vakanın (veya birkaç vakanın) derinlemesine incelenmesi, vaka çalışması hâline gelir (Creswell, 2013). Eğitim, sağlık, sosyal hizmet, yönetim ve diğer sosyal alanlar, tam da insanların gündelik hayatlarıyla ilgili oldukları için uygulamalı sosyal bilimler veya uygulama alanları olarak kabul edilir (Merriam & Tisdell, 2016).

1.1. Nitel Araştırmalarda Görüşme Yöntemi ve Uygulama Teknikleri

Nitel araştırmalarda en sık başvurulanan veri toplama araçları arasında yarı yapılandırılmış görüşme formları, odak grup görüşmeleri, katılımcı gözlem ve doküman analizi formları yer alır (Yıldırım & Şimşek, 2018). Bu araçlar vasıtasıyla araştırmacı, bireylerin kendi deneyimlerini, bakış açılarını ve yorumlama biçimlerini derinlemesine keşfetme olanağı elde eder. Bu durum, özellikle sosyal bilimlerde bireyin sosyal çevresiyle kurduğu ilişkiler ve bu ilişkilerin, taşıdığı anlamlar üzerinden çözülmesi açısından büyük önem taşır. Veri toplamının birincil biçimi genellikle araştırmacının katılımcılardan elde ettiği verileri, ortaya çıkan teori hakkındaki fikirlerle sürekli olarak karşılaştırdığı görüşmelerdir. Görüşme, temel bir sorgulama biçimidir (Seidman, 2006; Creswell, 2013). Çalışılan olguyu anlamak için gerekli nitel verilerin temel kaynağını oluşturan bu yöntem, soruların ve sıralarının önceden belirlendiği yapılandırılmış görüşmelerden, hiçbir unsurun önceden planlanmadığı, yapılandırılmamış görüşmelere kadar farklı biçimlerde uygulanabilir. Bunlar arasından en yaygın olanı, bir dizi soru ve keşfedilecek konu tarafından yönlendirilen, cümlelerin ve soruların sırasının tam anlamıyla önceden belirlenmediği yarı yapılandırılmış görüşmedir (Merriam & Tisdell, 2016).

Görüşme verilerinin kaydedilmesinde temel olarak kayıt cihazı kullanma ve not alma yöntemleri izlenir. Kayıt cihazı kullanımı, araştırmacının not alma sorununu ortadan kaldırarak dinleme ve soru sorma işlevlerine odaklanmasını ve böylece daha verimli bir görüşme yapmasını sağlar. Kayıt cihazının mümkün olmadığı durumlarda başvurulanan not alma yöntemi, yapılan görüşmeler daha yavaş ilerleyeceği için notların eksik alınmasına ve görüşme süresinin uzamasından dolayı zaman kaybına neden olabilir (Yıldırım & Şimşek, 2018). Not alma yöntemindeki bu dezavantajlar sebebiyle görüşme verilerini kaydetmenin en yaygın şekli ses kaydı kullanımıdır. Bu uygulama, söylenen her şeyin analiz için korunmasını sağlar. Aynı zamanda görüşme sırasında soru sorma tekniğinin geliştirilebilmesi için görüşmeciyi dinleme süresi kalmış olur. Bazen görüşmeler videoya kaydedilir. Bu yöntem, sözsüz davranışların da kaydedilmesini sağlar, ancak ses kaydı yapmaktan daha zahmetli bir yöntemdir (Merriam & Tisdell, 2016). Görüşmelerde ses kaydı alınması, birçok yönden faydalı bir uygulamadır. Katılımcıların sözlerinin korunması sayesinde araştırmacılar orijinal verilere sahip olur. Bu durum transkripsiyondaki belirsizlik durumlarında kaynağa dönüp doğruluğu kontrol etme imkânı verir. Röportaj materyallerini kötü kullanma iddiaları meydana gelirse araştırmacılar veriye karşı sorumluluklarını kanıtlamak veya kendi röportaj tekniklerini inceleyip geliştirmek amacıyla kayıtlardan yararlanabilirler (Seidman, 2006). Kayıt cihazı kullanmak, görüşmecinin yalnızca hafızasına güvenerek yakalayabileceğinden çok daha fazlasının kaydedilmesini sağlar. Konuşulanların kayıt altına alınması; söylenenlerin erişilebilir bir kaydının olması, görüşmenin sorumlu bir şekilde ele alınacağı konusunda güven vermesi gibi nedenlerle katılımcılara da fayda sağlar (Taylor, Bogdan & DeVault, 2016).

Günümüzde çeşitli boyut ve özelliklerde üretilen kayıt cihazları, özellikle sözlü anlatımların, halk müziği ürünlerinin ve geleneksel uygulamaların anlatım kısımlarının kaydedilmesinde pratiklik sağlamaktadır. Ses kaydının yanında video kaydı da veri toplama sürecinde belirli bir kolaylık sağlamakta ve özellikle kültürel aktarıma konu olan sözlü ifadelerin ve geleneksel anlatıların belgelenmesi açısından uzun vadeli bir arşiv işlevi görmektedir. Bu bağlamda sesli ya da görüntülü kayıt alınması, veri doğruluğu sağlamakla birlikte folklor, müzik, ritüel gibi canlı performansa dayalı kültürel olgulardaki beden dili, jest, mimik, tonlama, durak, ezgi gibi parçalarüstü birimlerin de analiz edilebilmesi bakımından önem taşımaktadır. Ancak bu araçların belli bir mesafeden sesi kaydetme özelliği nedeniyle, bir kaynak kişinin anlatımı veya icrası sırasında ortamdaki her türlü sesi de kaydedeceği unutulmamalıdır. Buna bağlı olarak gürültülü bir ortamda yapılacak kayıtlar sağlıklı olmaktan uzak kalabilir, istenmeyen seslerin kayda girmesi kaçınılmaz olabilir. Bu sebeple bir derlemeci, kullanacağı araçların teknik özellikleri hakkında bilgi sahibi olmalı ve ortam koşullarına göre yaka mikrofoni veya başka bir harici mikrofon kullanma yoluna gidebilmelidir (Ekici, 2019).

1.2. Görüşme Yönteminde Kayıt Kalitesi Sorunları ve Teknik İyileştirme Gereksinimi

Nitel araştırmalarda, özellikle alan temelli veri toplama süreçlerinde ses ve video kayıtları hem araştırma verilerinin güvenilirliğini artırmak hem de sözlü kültürün dokümantasyonunu sağlamak amacıyla sıklıkla kullanılan araçlardır. Fakat alan araştırmalarında ses ve video kayıtlarının kalitesini etkileyen unsurlar yalnızca cihaz teknolojisine değil, büyük ölçüde kaynak kişinin bireysel özelliklerine ve araştırma ortamının fiziksel koşullarına bağlıdır. Bu kayıt süreçleri çevresel ve mekânsal sorunlar bakımından gürültü ve ses bozulması, akustik ve ışık koşullarının elverişsizliği, kayıt cihazlarının teknik kısıtları, ani kesintiler ve bölünmeler; kaynak kişiye bağlı sorunlar bakımından ise yaşlılık ve fiziksel yetersizlikler, teknolojik güvensizlik ve kayıttan rahatsızlık duyma, anlamsal sapmalar, dilsel sorunlar ve psikolojik durumlar olarak sınıflandırılabilir.

Çevresel faktörler veri toplama sürecini zorlaştıran unsurlar arasında yer alır. Kırsal ya da açık alanlarda yapılan görüşmelerde rüzgâr, araç geçişi, hayvan sesleri gibi çevresel gürültüler; kapalı mekânlarda ise yetersiz aydınlatma, yankı gibi akustik sorunlar kayıt kalitesini olumsuz etkileyebilir. Araştırmacının bu koşullar üzerinde doğrudan kontrolü olmadığından kayıt sırasında ortaya çıkan ses bozulmaları veya görüntü kayıpları daha sonra yapılacak analizlerin güvenilirliğini azaltabilir. Genellikle, bir ses kasetinin kaydı çok zayıf olabilir veya ilgili konuşmaların daha sonra fark edilmeyen bir arka plan gürültüsü tarafından bastırıldığı görülebilir (Yin, 2011). Görüşme mekânı, kayıt kalitesi ve katılımcının rahatlığı açısından büyük önem taşır. Arka planda çok fazla gürültü olan bir mekân, kayıtların kalitesini olumsuz etkileyen sonuçlar ortaya çıkarabilir. Gürültüden dolayı sesleri duymak zorlaşabilir. Ayrıca, görüşülen kişinin halka açık bir mekânda samimi bir şekilde konuşabileceğinden emin olunması ve görüşme mekânının, katılımcının mümkün olduğunca rahat olabileceği bir yer olarak seçilmesi gerekir (Horvat vd., 2013). Alan araştırmalarında doğal ortamdan kaynaklanan sesler ortaya çıkabilir. Bunlar arasında ezan okunması, belediye hoparlöründen duyurular, köy odası veya kahvehane gibi kalabalık ortamlardaki sesler, hayvan sesleri, araya giren konuşmacılar, çocuk sesleri, telefon sesi ve rüzgâr sesi gibi unsurlar başta gelir. Doğal ortamlarda kayıt sırasında kaçınılması her zaman mümkün olmayan bu sesler, derlemelerde seyrek de olsa yer alabilmektedir (Demir, 2013).

Veri toplama sürecini ve kayıt kalitesini etkileyen diğer önemli unsur kaynak kişilerdir. Görüşme yapılan kişinin sağlığı, görüşme sırasındaki psikolojik durumu ve projenin katılımcısı olmaya yönelik niyetleri dâhil olmak üzere, bu tür bireysel durumlar elde edilen verilerin niteliğini etkileyebilir (Merriam & Tisdell, 2016). Özellikle alan araştırmalarında kaynak kişilerin yaşlılık, işitme veya konuşma güçlüğü gibi fiziksel yetersizlikleri, psiko-sosyal gerilim hâlleri ya da teknolojik araçlara karşı güvensizlikleri ses ve video kaydının niteliğini doğrudan belirleyebilmektedir. Örneğin, ileri yaşta bir kaynak kişinin çok düşük tonda konuşması, çene ve diş yapısındaki zafiyet, işitme yetersizliği vb. durumlar kelimelerin anlaşılmasını ve beraberinde kayıtların çözümlenebilmesini güçleştirebilmektedir. Bunun yanında, bazı kişiler ses veya görüntü kaydı alınmasını geleneksel değerler ya da kişisel mahremiyet algısı nedeniyle rahatsız edici bulabilmekte ve bu durum doğal ifade akışını kesintiye uğratabilmektedir.

Alan araştırması, derleme gibi adlarla anılan görüşmelerde veri toplama sürecinde ortaya çıkan sorunlara ilişkin çeşitli öneriler getirilmiştir. Bunlar arasında; kamera için en iyi konumun belirlenmesi (Creswell, 2013), cihazların görüş alanının dışına yerleştirilmesi (Taylor, Bogdan & DeVault, 2016), her görüşmeden önce ekipmanın düzgün çalıştığından emin olunması (Taylor, Bogdan & DeVault, 2016), kayıt ekipmanlarının çeşitlendirilmesi ve kayıt mesafesinin doğru ayarlanması (Ekici, 2019) gibi çözüm önerileri sayılabilir. Ancak araştırmacının elinde olmayan çevresel ve kişisel nedenlerden kaynaklanan istenmeyen seslerin tam olarak önüne geçilememektedir.

Gelişen teknoloji sayesinde, ses ve görüntü kayıtları kaliteli dijital formatlara aktarılabilir. Son geliştirilen müzik programları sayesinde aktarım esnasında istenmeyen seslerin ayıklanması mümkün olmakta, böylece kayıtların dinlenmesi, yedeklenmesi kolaylaşmakta ve fiziksel yıpranma riski ortadan kalkmaktadır (Öçalan, 2006). Araştırmacıların bal mumu silindir gibi tarihsel ortamlardan günümüzün dijital formatlarına kadar topladıkları tüm kayıtları, uygun yöntemlerle işleyerek belirli ölçüde temizlemek; bu sayede klima uğultusu, projektör hışırtısı veya rüzgâr gürültüsü gibi arka plan sesleri içeren kayıtlar üzerinde duyulabilir iyileştirmeler sağlamak mümkündür.

Alan araştırmalarında ses kayıt kalitesini etkileyen ve yukarıda belirtilen sorunlar, nitel verilerin analizini güçleştirmektedir. Söz konusu sorunların bir bölümü kayıt öncesinde alınacak önlemlerle

kısmen azaltılabilse de kontrol edilemeyen çevresel koşullar ve kaynak kişiye bağlı değişkenler, araştırmacı için kayıt sonrası teknik müdahale ihtiyacı doğurmaktadır. Araştırmacıların büyük çoğunluğu ses işleme konusunda teknik bir altyapıya sahip değildir. Mevcut ticari yazılımlar ise maliyetli ve yeniden üretilebilirlik açısından sınırlı kalmaktadır. Bu durum, nitel araştırmacılara yönelik açık kaynaklı, erişilebilir ve teknik uzmanlık gerektirmeyen ses iyileştirme araçlarına duyulan ihtiyacı ortaya koymaktadır. Buradan hareketle bu çalışmada, görüşme kayıtlarında karşılaşılan gürültü ve bozulma sorunlarını sistematik biçimde ele alan, Python tabanlı dijital sinyal işleme ve yapay zekâ destekli bir hibrit iyileştirme hattı tasarlanmıştır. Böylece veri kalitesinin artırılması ve transkripsiyon doğruluğunun iyileştirilmesi hedeflenmektedir.

2. YÖNTEM

Bu çalışmada, araştırmalar için önemli veriler içeren ses ve video kayıtlarındaki istenmeyen seslerin giderilmesini sağlayacak pratik öneriler ve teknik çözümler sunulacaktır. Bu bağlamda, tarihî/analog kayıtlar yalnızca örnek düzeyinde anılacak; restorasyonda başvuru alan temel ilkeler (ör. bant hışırtısı ve mekanik uğultunun azaltılması, konuşma anlaşılabilirliğinin korunması) kısaca özetlenecektir. Çalışmanın ana odağı ise çağdaş saha ve laboratuvar kayıtlarında pratik iyileştirme adımlarının uygulanabilirliğidir. Problemlere modern ve otomatize edilmiş bir yaklaşımla odaklanılarak istenmeyen sesleri düzeltmek için ses kayıt programlarında uygulanacak yöntemler, yapay zekâ destekli çözümler ve yazarlar tarafından oluşturulan yazılım kodları paylaşılacak ve bu uygulamaların nasıl kullanılacağına dair açıklamalar yer alacaktır. Belirtilen amaç doğrultusunda, araştırmacıların görüşmelerden elde ettikleri kayıtlar için enerji ve spektral düzlük tabanlı adaptif eşikleme, uğultu ve harmonik filtreleme, tam-adaptif spektral gürültü azaltma, konuşmayı ön plana çıkarma ve klik/plosive temizliği gibi aşamaları otomatik biçimde uygulayan Python betikleri oluşturulmuştur. Bu süreçte, yukarıda bahsedilen türde arka plan gürültüsü içeren görüşme kayıtları kullanılarak kod defteri test edilmiş ve bu modüller tek bir otomatik işleme zinciri içinde bütünleştirilmiştir. Kod defterinden ve araştırmacı tarafından manuel olarak oluşturulan referans transkripsiyonlar, kelime ve karakter doğruluğunu ölçen bir Python betiğiyle karşılaştırmalı olarak değerlendirilmiştir. Araştırmacıların farklı ortamlarda elde ettikleri kayıtları daha kolay, tutarlı ve tekrarlanabilir biçimde analiz edebilmelerini sağlamak amacıyla, çalışmanın sonraki bölümünde geliştirilen çevrimiçi arayüz ve adaptif Python tabanlı işlem adımları ayrıntılı biçimde ele alınmıştır.

Bu çalışmada, geliştirilen işleme hattının performansını ölçmek amacıyla, nitel saha araştırmalarının karakteristik gürültü koşullarını ve akustik zorluklarını yansıtan bütünlük bir kayıt senaryosu tasarlanmıştır. Bu bağlamda araştırmacılar tarafından seslendirilen standart metinlerden elde edilen veri kümesi, kayıt sonrası işleme stratejilerinin gürültü azaltma performansı ile transkripsiyon doğruluğu üzerindeki etkisini sistematik biçimde değerlendirmek için kullanılmıştır. Bu kontrollü tasarım, önerilen iyileştirme zincirinin gürültü azaltma ve transkripsiyon doğruluğu üzerindeki etkisinin gerçek saha koşullarına benzer karmaşıklıkta bir düzlemde sistematik olarak değerlendirilmesine olanak tanımıştır.

2.1. Etik Kurul Onayı

Bu araştırmanın kavramsal çerçevesinin hazırlanması, veri toplama araçlarının uygulanması, verilerin toplanması, verilerin analizi ve yorumlanması aşamalarının tamamında etik kurallara uygun hareket edilmiştir. Karşılaşılabilecek tüm etik ihlallerde ANKAD Yayın Kurulunun hiçbir sorumluluğu bulunmamaktadır. Tüm sorumluluk yazarlara aittir. Bu çalışmanın ANKAD dışında herhangi bir akademik yayın ortamına değerlendirme için gönderilmediği taahhüt edilmektedir. Yapılan bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbirisi gerçekleştirilmemiştir. Görüşmelerden elde edilen ses kayıtlarının incelenmesinde nitel araştırma yöntemlerinden "Doküman Analizi" tekniği uygulandığı için etik kurul kararı gerektirmemektedir.

3. GÜRÜLTÜ AZALTIMI VE KONUŞMA İYİLEŞTİRMeye YÖNELİK ÖNERİLER

3.1. Kayıt Sürecine İlişkin Yöntemsel Öneriler

Araştırma ortamlarının doğası gereği değişkenlik göstermesi, araştırmacının esnek, önleyici ve bağlama duyarlı hazırlıklar yapmasını gerektirmektedir. Açık alanlarda yürütülen görüşme ve ses kayıt

süreçlerinde rüzgâr, trafik veya insan kalabalığı gibi çevresel sesler, ses kaydının bütünlüğünü bozabilmektedir. Bu doğrultuda gürültünün ölçülmesi, filtrelenmesi veya donanım ile en aza indirgenmesi önemlidir. Kapalı alanlarda ise yankı (reverberation), akustik yansımalar ve mikrofon-donanım yerleşimi gibi faktörler, akustik çevrenin konforu, sesin netliğini ve kalitesini olumlu ya da olumsuz yönde etkileyebilmektedir. Yankı süresi (reverberation time), iç-dış kaynaklı seslerin katkısı ve sesin mekânlardaki sürekliliği gibi mekân akustiği öğeleri ile mikrofon karakteristiklerinin birlikte değerlendirilmesi bu açıdan önemlidir (Bottalico vd., 2020; Cui vd., 2021; Eldwaik & F. Li, 2018; Torresin vd., 2020). Dolayısıyla öncelikle kayda elverişli ortamın oluşturulması ve kayıt zincirinin ilk halkası olan mikrofonun da açık ya da kapalı kayıt alan özelliklerine göre uygun şekilde seçilmesi gerekmektedir.

Bu tür durumlarda veri kalitesini korumak için öncelikle açık alanlarda, kayıt süresince ortaya çıkabilen düşük frekanslı çevresel gürültüleri azaltmak için mikrofon ya da ses kartı/mikser üzerindeki yüksek geçiş (high-pass) filtrelerinin aktif edilmesi etkili bir yöntemdir. Özellikle 80 Hz ve altındaki frekansların bastırılması trafik, rüzgâr veya titreşim kaynaklı uğultuları azaltarak konuşma odaklı net kayıtlar elde etmeyi sağlamaktadır. Kapalı ortamlarda ise yankı ve rezonans sorunlarını gidermek amacıyla mekâna uygun akustik düzenlemeler yapılmalı; gerekirse kalıcı ya da taşınabilir ses difüzörleri, bas tuzakları ve akustik paneller kullanılarak yansıma ve mekânsal bozulmalar kontrol altına alınmalıdır (Aljumaily & Abdulmohsin, 2024; MacCallum vd., 2011).

Ses kayıtlarında seçilecek mikrofon türü önemlidir. Kondenser mikrofonlar; geniş frekans tepkisi, yüksek duyarlılık ve düşük öz-gürültüleri nedeniyle ayrıntı yakalamada diğer türlere göre daha detaylıdır ve tercih edilir (Eargle, 2012; Titze & Winholtz, 1993). Çevresel gürültülerin ve mekanik titreşimlerin etkisini azaltmak için mikrofon ile pop filtre, rüzgâr koruyucu ve titreşim engelleyici (shock-mount) gibi destekleyici ekipmanların kullanımı gerekmektedir. Bu ekipmanların saha koşullarında kayıt kalitesini iyileştirdiği görülmüştür (Eargle, 2012; Özyıldırım, 2010; Titze & Winholtz, 1993). Kayıt sürecinde mikrofonun konumlandırılması, konuşma/sözlü anlatı kayıtlarında mikrofonun anlatıcının ağzına doğru hafif açılı ve yaklaşık 15-30 cm mesafede yerleştirilmesi arka plan gürültüsünü ve patlamalı ünsüz kaynaklı bozulmaları azaltarak doğrudanlığı artırmaktadır. Ayrıca mikrofon türüne (kondenser, dinamik) ve kardiyotid/hyper-kardiyotid gibi yönelim karakteristiğine bağlı olarak kaynak-mikrofon mesafesinin ve ekseninin (on/off-axis) doğru seçiminin, netlik üzerinde ölçülebilir farklar yarattığı bilinmektedir (Corbett, 2014; Konuma & Asakura, 2023; Titze & Winholtz, 1993). Dolayısıyla hem hedef ses ön plana çıkarılabilmekte hem de çevresel yansımaların kayıt üzerindeki etkisi azaltılabilmektedir.

Araştırmacılar, cep telefonlarında yer alan ses kayıt uygulamalarını ya da dizüstü bilgisayarlarda bulunan dijital ses düzenleme yazılımlarını kullanarak kayıt süreçlerinde esneklik ve pratiklik kazanabilmektedir. Ses kayıtları alınırken donanım ve yazılım bileşenlerinin birlikte kullanılması kritik bir unsur olduğundan sürecin verimliliği için bu bileşenlerin önceden planlanması gerekmektedir. Konuşma kayıtlarının anlaşılabilirliği ve çözünürlüğü; kullanılan ekipman, sinyalin teknik özellikleri, kayıt ortamının akustik yapısı, ses düzeyi ve kayıt sonrası uygulanan düzenleme işlemlerine bağlı olarak değişmektedir (Rogers vd., 2020; Vergili, 2017).

3.2. Kayıt Sonrasına İlişkin Yöntemsel Öneriler

Kayıt aşamasından sonra araştırmacılar, elde edilen verilerdeki istenmeyen gürültü ve bozulmaları iZotope RX gibi profesyonel yazılımlar ve hazır algoritmalar aracılığıyla azaltabilmektedir. Ancak bu çalışma kapsamında, ticari yazılımlara bağımlılığı azaltmak ve süreçleri tekrar üretilebilir, şeffaf ve özelleştirilebilir hâle getirmek amacıyla Python programlama dili kullanılarak çevrimiçi erişilebilen bir web arayüzü tasarlanmış ve gürültü azaltma ile konuşma iyileştirme adımları tek bir otomatik işleme zinciri içinde bütünleştirilmiştir.

Ses dosyaları üzerinde gerçekleştirilen gürültü ve benzeri istenmeyen seslerin giderilmesi sürecinde iZotope RX yazılımı, öne çıkan bir araçtır. RX'in arayüzünde yer alan *De-click*, *De-hum*, *Spectral De-noise*, *Breath Control* ve *Spectral Repair* gibi düzenleme algoritmaları, kayıtların işlenmesi ve düzenlenmesi aşamalarında kullanılmaktadır. Bu kapsamda filtreler ve efektler aracılığıyla sesin daha anlaşılır hâle getirilmesi mümkündür. Konuşmacıdan kaynaklanan sibilans sesler, ağız hareketlerine bağlı gürültüler ya da aşırı nefes sesleri RX yazılımının *De-ess*, *Mouth De-click* ve *Breath Control*

modülleriyle giderilebilirken kayıt ortamına özgü ambiyans gürültüleri ve çevresel istenmeyen sesler ise *Spectral De-noise* ve *Spectral Repair* araçlarıyla en aza indirilebilmektedir (Erdal, 2023).

İyileştirme işlemlerinin temel amacı, ses sinyalinde minimum bozulma yaratarak kaliteyi artırmak ve çevresel gürültüleri mümkün olduğunca düşük seviyelere indirmektir. Düzenleme sürecinde ses dosyalarının kulaklıkla dinlenmesi, daha doğru ve ayrıntılı bir değerlendirme yapılmasına olanak tanımaktadır. RX yazılımı örneğinden hareketle, kayıtların düzenlenmesinde izlenebilecek aşamalar sistematik bir yaklaşım çerçevesinde uygulanabilir (Erdal, 2023). Rogers vd. (2020), temel düzenleme işlemleri (örneğin kesme, kopyalama ve yapıştırma) için ise Audacity veya Adobe Audition gibi ses düzenleme yazılımlarından da yararlanılabildiğini ifade etmişlerdir. Bu programların ve kayıt sonrası yöntemlerin doğru şekilde uygulanabilmesi ve araştırmacıların bu işlemleri yerine getirebilmesi için temel bilgi düzeyine, ayrıca uygun ekipmanlara sahip olması gerekmektedir. Bununla birlikte, bu çalışma kapsamında söz konusu işlemler ticari yazılım ortamı yerine Python tabanlı, tekrar üretilebilir ve özelleştirilebilir bir işlem hattı üzerinden gerçekleştirilmiştir.

Bu çalışma kapsamında geliştirilen betik, araştırmacıların kendi saha/görüşme kayıtlarını yükleyerek aynı işleme zincirini tekrarlanabilir biçimde çalıştırabilecekleri uygulanabilir bir araç olarak tasarlanmıştır. Arayüz üzerinden dosya yükleme, ön işleme çıktılarının üretilmesi, ses temizleme, segmentasyon ve otomatik Türkçe transkripsiyon adımlarının tek bir akış içinde yürütülmesi hedeflenmiş böylece hem standartlaştırılmış çıktı üretimi hem de süreç izlenebilirliği amaçlanmıştır. Bununla birlikte sistem, farklı donanım koşullarına uyarlanabilmesi ve yeni model/algoritmalarla güncellenebilmesi için modüler biçimde kurgulanmıştır. Dolayısıyla mevcut sürüm “çalışır bir prototip” olarak değerlendirilirken geliştirmeye açık bir tasarım çerçevesi korunmuştur. Araştırmacılar tarafından kullanılabilir bir araç şeklinde ulaşılabilmesi için *huggingface* sitesinde yayınlanmıştır.⁴

3.2.1. Dijital Sinyal İşleme (DSP) ve Ön İşleme

3.2.1.1. Kütüphanelerin Kurulumu

Bu bölümde, ses işleme ve otomatik transkripsiyon süreçlerini destekleyen dijital ön işleme zincirinin yazılım altyapısı kurulmuş ve süreçlerin otomatikleştirilebilir biçimde yürütülmesi amaçlanmıştır. Bu aşamada, ses işleme ve otomatik transkripsiyon süreçlerinin yürütülebilmesi için gerekli yazılım altyapısı yapılandırılmıştır. Betik, öncelikle ses işleme için sistem düzeyinde gerekli olan *ffmpeg* bağımlılığını kurmakta, ardından bilimsel hesaplama ve sinyal işleme kütüphanelerini (*numpy*, *scipy*, *librosa*, *soundfile*) yüklemektedir. Bu doğrultuda araştırmacılar veya ses analizi yapan kullanıcılar farklı formatlardaki (wav, mp3, FLAC, m4A vb.) ses kayıtlarını doğrudan yükleyip önizleyebilmelerini sağlayan bir giriş aşamasını oluşturmaktadır.

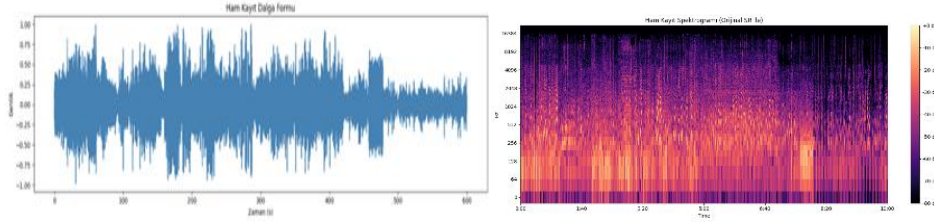
Gürültü azaltma ve konuşma tespiti gibi işlemler için *noisereduce* ve *webrtcvad*; efekt ve filtreleme süreçleri için *pedalboard*; veri görselleştirme için *matplotlib*; süreç takibi için *tqdm*; dosya işleme ve dönüştürme işlemleri için *pydub*; etkileşimli arayüz geliştirme için *gradio* kütüphaneleri kurulmaktadır. Bununla birlikte, derin öğrenme tabanlı kaynak ayrıştırma işlemleri için Demucs modeli (v4.0.1) yüklenmiş ve kurulum sonrası doğrulama testi gerçekleştirilerek çalışma ortamının kararlılığı kontrol edilmiştir. Bu yapılandırma adımı, ilerleyen aşamalarda uygulanacak gürültü azaltma, filtreleme, spektral analiz, diyalog izolasyonu ve otomatik transkripsiyon işlemleri için tekrar üretilebilir bir çalışma ortamı sağlamakta; farklı kaynaklardan elde edilen ses kayıtlarının tek bir işlem zinciri içinde standart ve karşılaştırılabilir bir formata dönüştürülmesine olanak tanımaktadır.

3.2.1.2. Ses Dosyası Yükleme ve Dinleme, Ham Kayıt Spektrogramı

Bu aşamada ham ses kaydı herhangi bir gürültü azaltma veya filtreleme uygulanmadan yüklenmiş; sinyalin spektral yapısı ve dinamik aralığı doğrudan incelenmiştir. Ham verinin zaman alanındaki dalga formu ve zaman-frekans düzlemindeki dağılımı analiz edilerek ilerleyen safhalarda uygulanacak filtreleme, gürültü azaltma ve konuşmayı ön plana çıkarma işlemlerine ilişkin kararların daha bilinçli biçimde verilmesi amaçlanmıştır. Özellikle formant bölgeleri, olası uğultu frekansları ve sessiz segmentlerin dağılımı bu görselleştirme aracılığıyla belirlenmiştir. Spektrogram üretiminde sinyalin Kısa Süreli Fourier Dönüşümü (STFT) alınmış ve genlik değerleri dB ölçeğine dönüştürülerek zaman

⁴ <https://huggingface.co/spaces/kyerdal/Transcription>

içinde değişen frekans bileşenleri görünür kılınmıştır. STFT temelli analizler, konuşma sinyalinin frekans içeriğini ve gürültü bileşenlerini belirlemede yaygın olarak kullanılan bir yöntemdir (Özkurt, 2025).



Şekil 1. Ham ses kaydı Fourier dönüşüm ve spektrogramı

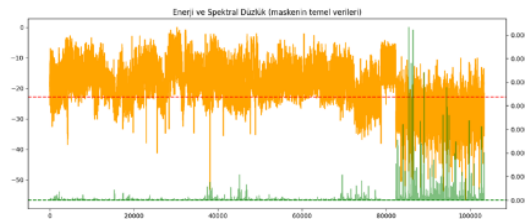
Kısa Süreli Fourier Dönüşümü, insan işitme algısına daha yakın analizler elde edebilmek amacıyla bazı çalışmalarda A-weighting gibi algısal ağırlıklandırma filtreleriyle birlikte kullanılabilir. Benzer biçimde, kayıtlarda sıkça rastlanan şebeke uğultusu (hum), klik ve nefes sesleri gibi spesifik bozulmaların giderilmesi için dar bantlı filtreler, spektral onarım algoritmaları ve yüksek geçiş filtreleri gibi çeşitli teknikler literatürde önerilmektedir (Erdal, 2023; Özkurt, 2025; Wang & Yu, 2025). Bu tür yöntemlerin programlanabilir ortamlarda uygulanabilmesi, ses işleme süreçlerinin otomatize edilmesine olanak tanımaktadır.

Bu aşamada, ham ses kaydı herhangi bir gürültü azaltma veya filtreleme uygulanmadan yüklenmiş; sinyalin zaman alanındaki dalga formu ve zaman-frekans düzlemindeki spektrogramı üretilerek görsel inceleme için arşivlenmiştir. Kod, ses dosyasını orijinal örnekleme oranını koruyarak okumuş ve mono formata indirgemıştır. Ham kayıt ayrıca yeniden kullanılabilirlik amacıyla WAV biçiminde kaydedilmiş; dalga formu ve spektrogram görselleri ise sırasıyla PNG formatında dışa aktarılmıştır.

Spektrogram üretiminde, sinyalin STFT alınmıştır. Genlik değerleri dB ölçeğine çevrilerek zaman içinde değişen frekans bileşenleri görünür kılınmıştır. Bu görselleştirme, ilerleyen aşamalarda uygulanacak gürültü azaltma, uğultu temizliği ve konuşma ön-plana çıkarma adımları için başlangıç düzeyinde bir akustik referans sağlamaktadır.

3.2.1.3. Enerji ve Spektral Düzlük ile Gürültü Karelerinin Tespiti

Bu aşamada, ses sinyalinin kare (frame) düzeyindeki enerji yoğunluğu ve spektral düzlük (spectral flatness) değerleri birlikte değerlendirilerek kayıttaki olası gürültü bölgeleri otomatik biçimde tespit edilmiştir. Enerjisi görece düşük ve spektral olarak düz (tonal içeriği zayıf) alanlar, arka plan gürültüsünün baskın olduğu bölümler olarak etiketlenmiş ve bir gürültü maskesi oluşturulmuştur. Adaptif eşikleme yaklaşımında sabit bir eşik değeri yerine, kayıt içindeki enerji ve spektral düzlük dağılımlarının yüzdelik değerleri temel alınarak veri güdümlü eşikler belirlenmiştir. Böylece eşikler, her bir kaydın genel enerji seviyesine bağlı olarak dinamik biçimde uyarlanmıştır.



Şekil 2. Gürültü karelerinin tespiti

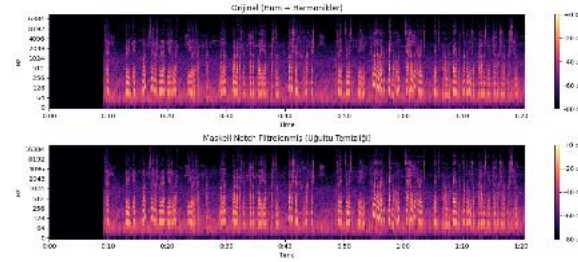
Elde edilen gürültü maskesi, geçişlerin daha doğal olması amacıyla Gaussian filtreleme ile yumuşatılmıştır. Bu pürüzsüzleştirilmiş maske önce kare düzeyinde tanımlanmış, ardından örnek (sample) düzeyine genişletilerek zaman alanında uygulanabilir hâle getirilmiş ve sonraki gürültü azaltma aşamalarında kullanılmak üzere dosya çıktısı olarak kaydedilmiştir (bk. Şekil 2).

Spektral düzlük, bir sinyalin spektral içeriğinin gürültüye mi yoksa tonal bir yapıya mı daha yakın olduğunu belirlemede yaygın biçimde kullanılan bir ölçüttür. Benzer biçimde, çok bantlı spektral çıkarım (MBSS) yaklaşımlarında da bant bazlı SNR tahminlerine dayalı adaptif faktörler kullanılarak düşük SNR koşullarında gürültü bölgelerinin daha hassas tespit edilebildiği gösterilmiştir (Bhattacharya

& Tangsangiumvisai, 2017; Segerstrom vd., 2022). Bu bağlamda enerji ve spektral düzlük temelli maskelenme, sonraki tam-adaptif gürültü azaltma adımı için temel bir karar mekanizması oluşturmaktadır.

3.2.1.4. Eşiklemeye dayalı (maskelenmiş) çentik filtreleme ile uğultu azaltma

Ses kaydındaki uğultu (hum) ve harmonikleri otomatik olarak tespit edilmiş olup yalnızca gürültü maskesiyle belirlenen bölgelerde temizlenmektedir. Böylece, konuşma içeriği korunurken gereksiz frekans kayıpları önlenmiştir. Bu uğultu benzeri gürültüler genellikle elektrik kaynaklı (50/60 Hz) parazitlerden veya kayıt ortamındaki zemin titreşimlerinden kaynaklanmaktadır. Bu çalışmada ise temel uğultu frekansı (40–200 Hz) aralığında spektral tepe analiziyle otomatik olarak belirlenmiştir. Maskeli notch filtreleme yaklaşımı ile bu parazitleri hedef alarak sesin doğal tınısı korunmuştur. Bununla birlikte hem spektrogram karşılaştırması hem de araştırmacı tarafından yapılan dinleme değerlendirmesi üzerinden temiz ve dengeli bir taban ses elde edilmiştir.

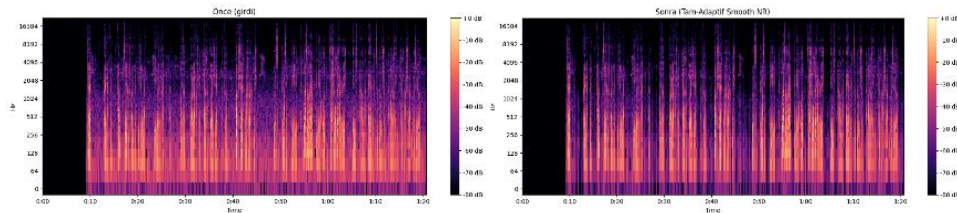


Şekil 3. Uğultu (Hum) ve gürültü maskesi entegrasyonu öncesi ve sonrası

Yaygın olarak 50/60 Hz uğultunun bastırılması için dar bantlı çentik (notch) filtreler kullanılmaktadır. Bu tür filtrelerin, *scipy.signal.iirnotch* gibi Python kütüphaneleriyle kolayca uygulanabilir olması, multidisipliner projelerde pratik avantajlar sunmaktadır. Dar bantlı çentik filtrelerin konuşma sinyalinin ana spektral içeriğine minimum müdahale ile uğultuyu bastırabildiği gösterilmiştir. Ayrıca uğultu ve harmonikler, önceden oluşturulan gürültü maskesiyle sınırlandırılmış bölgelerde daha etkili bir şekilde temizlenebilmektedir (Afifah vd., 2018; Retdian & Afifah, 2020). Bu aşamada elde edilen sinyalin nitelik bilgisine dayalı olarak tam-adaptif yumuşak gürültü azaltma işlemine geçilmiştir.

3.2.1.5. Sinyal Bilgisine Dayalı Tam-Adaptif Spektral Gürültü Azaltma

Sinyalin enerji düzeyi, spektral merkez değeri ve tahmini sinyal-gürültü oranı (SNR) her bir kare için hesaplanarak adaptif bir gürültü azaltma süreci yürütülmüştür. Model, sinyalin enerji düzeyine ve tahmini (SNR) değerine göre konuşma ve sessizlik bölgelerini dinamik biçimde ayırmıştır. Her bir kareye özgü değişken gürültü azaltma oranları uygulanarak sesin doğal tınısı korunmuş ve bu sayede düşük enerjili konuşma bileşenlerinin hatalı biçimde bastırılmasına bağlı olarak oluşabilecek aşırı baskılama (over-suppression) veya boğukluk (timbre degradation) türü bozulmaların önüne geçilmiştir.

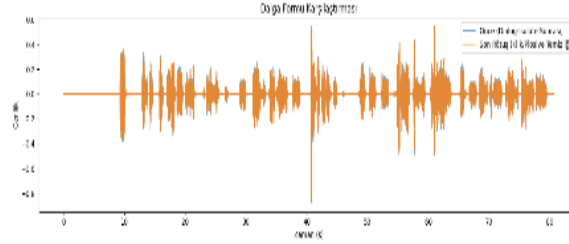


Şekil 4. Sinyal bilgisine dayalı tam-adaptif spektral gürültü azaltma öncesi ve sonrası

Tam-adaptif spektral gürültü azaltma yaklaşımı, özellikle konuşma sinyallerinde sinyal-gürültü oranına (SNR) ve spektral merkeze göre dinamik olarak ayarlanabilen tekniklerle daha etkili hâle getirilmiştir. Roy & Paliwal (2021), konuşma ve sessizlik karelerinin ayrılması için spektral düzlük tabanlı adaptif eşikleme yaklaşımında sessiz karelerde doğrudan gürültü tahmini yapılırken, konuşma karelerinde ileri düzey türev tabanlı filtreleme yöntemleri kullanılarak konuşma bileşenlerinin netliğinin korunduğunu göstermiştir. Bu çalışmada da kareler arasında yumuşak geçiş (temporal smoothing) uygulanarak işlem sonrası sesin sürekliliği korunmuştur. Böylece işlem sonrası sinyalde spektral süreklilik bozulmadan gürültü azaltma gerçekleştirilmiş ve akustik denge korunmuştur (bk. Şekil 4).

3.2.1.6. Konuşmayı Ön Plana Çıkarma

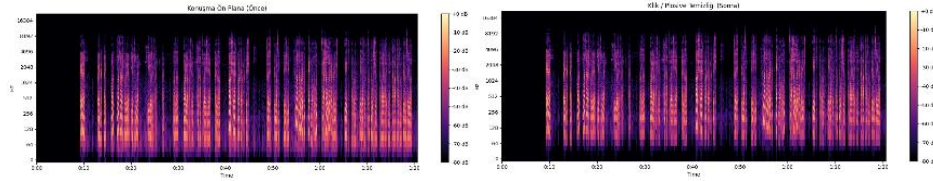
Gürültü azaltma işlemlerinin ardından, konuşma frekans bantlarının göreceli olarak öne çıkarılması amacıyla Dialog-Isolate yaklaşımı uygulanmıştır. Bu aşamada adaptif kazanç şekillendirme (adaptive gain shaping) gerçekleştirilmiş ve özellikle 2.8 kHz ile 6.5 kHz bantlarında netlik artırıcı ekolayzır (Presence EQ) uygulanmıştır. Böylece düşük frekanslardaki gereksiz enerji bastırılmış, konuşma açısından kritik üst-orta frekans bileşenleri korunmuştur. Geçiş sürekliliğini sağlamak amacıyla alçak geçiren (low-pass) yumuşatma filtresi uygulanmış ve ani spektral değişimlerin etkisi azaltılmıştır. Bu işlemler sonucunda elde edilen çıktı (y_{expanded}), arka plan bileşenleri göreceli olarak bastırılmış ve konuşma içeriği spektral olarak daha belirgin hâle getirilmiştir (bk. Şekil 5).



Şekil 5. Kazanç biçimlendirme ve gürültü azaltımı sonrası konuşma bileşenlerinin zaman alanı görünümü

Benzer biçimde (Nian vd., 2021), konuşma ön işleme aşamasında gelişmiş minima kontrollü yinelemeli ortalama (IMCRA) ile derin öğrenme tabanlı adaptif maskeleme tekniklerini birleştirerek çerçeve bazında gürültü takibi ve spektral kapılama uygulamışlardır. Söz konusu yaklaşımda konuşma aktivitesi tespiti (VAD) ile birlikte konuşma sırasında daha sınırlı, sessizlik karelerinde ise daha yüksek düzeyde gürültü azaltma uygulanmıştır. Bu tür tam-adaptif yaklaşımlar, Python tabanlı kütüphanelerle entegre biçimde uygulanabilmektedir. Spektral şekillendirme ve dinamik aralık düzenleme tekniklerinin, başlangıç aşamasında uygulanan gürültü azaltma işlemleriyle birlikte kullanıldığında konuşma anlaşılabilirliğini artırdığı literatürde gösterilmiştir. Griffin vd. (2015), bu tür sistemlerin kontrollü koşullarda yaklaşık 10 dB düzeyinde SNR kazanımına eşdeğer algısal iyileşme sağlayabildiğini belirtmiştir.

Bu çalışmada da işlem sonrası spektrogram karşılaştırmalarında 80 Hz altındaki düşük frekans enerjisinin azaldığı, buna karşılık 2–7 kHz aralığındaki konuşma açısından kritik bantların korunduğu gözlemlenmiştir (bk. Şekil 6). Adaptif kazanç şekillendirme ve üst-orta frekans ekolayzır uygulamaları sonucunda konuşma bileşenlerinin arka plan enerjisine göreceli üstünlüğü artırılmıştır.

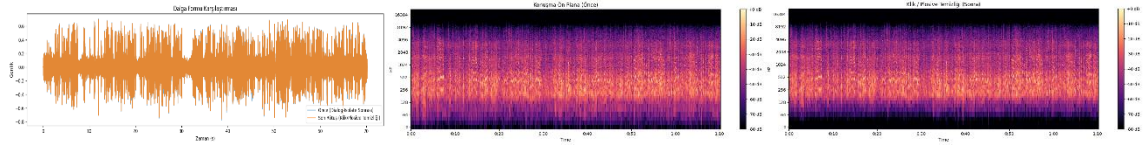


Şekil 6. Adaptif spektral gürültü azaltma öncesi ve sonrası konuşma sinyalinin spektrogram görüntüleri

3.2.1.7. Klik ve Plosive Temizliği (Yumuşak Rötüş Aşaması)

Kayıtlarda ani enerji sıçramaları (klik) ve patlayıcı ünsüz kaynaklı düşük frekans patlamaları (plosive) türündeki geçici bozulmalar, sinyalin temel spektral yapısı korunarak giderilmiştir. Klik bileşenlerinin bastırılmasında yerel çizgisel yumuşatma yaklaşımı uygulanmıştır. Plosive kaynaklı düşük frekans yoğunlaşmaları hafif bir yüksek geçiren (high-pass) filtre ile azaltılmıştır. Bu işlemler sonucunda geçici enerji sıçramaları kontrol altına alınmış ve elde edilen çıktı (y_{final}) spektral süreklilik açısından daha dengeli hâle getirilmiştir.

Klik ve plosive tespitinde türev (diferansiyel) tabanlı ani değişim analizi kullanılmıştır. Konuşma karelerinde sinyalin zamansal ani genlik artışları belirlenmiş ve bu tür bozulmalar izole edilmiştir. Düşük frekanslı plosive bileşenlerinin azaltılmasında Butterworth tabanlı IIR yüksek geçiren filtreler sıfır fazlı filtefilt uygulaması ile kullanılmış ve faz bozulmaları önlenmiştir (Magsi vd., 2018; Roy & Paliwal, 2021).



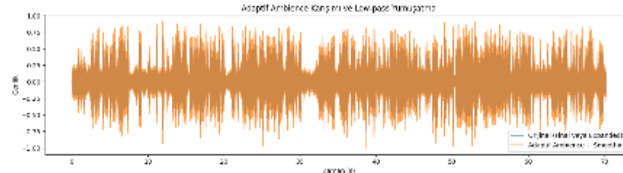
Şekil 7. Klik ve plosive temizliği öncesi ve sonrası

Önceki aşamalarda elde edilen enerji ve spektral düzlük temelli göstergeler, bu adımda klik ve plosive bölgelerinin belirlenmesinde tamamlayıcı rol oynamıştır. Çok bantlı spektral çıkarım yaklaşımlarında bant bazlı SNR'ye dayalı adaptif faktörlerin konuşma kalitesinin korunmasında avantaj sağladığı literatürde gösterilmiştir (Bhattacharya & Tangsangiumvisai, 2017). Roy & Paliwal (2021), da spektral düzlük tabanlı adaptif eşikleme ile konuşma ve gürültü bölgelerinin ayrımının klik ve plosive tespiti için temel oluşturduğunu belirtmiştir.

Bu aşamada ses dosyası üzerinde gerçekleştirilen işlemlerden sonra dalga formu analizinde (bk. Şekil 7), ani enerji sıçramalarının çizgisel yumuşatma yöntemiyle bastırıldığını; spektrogram karşılaştırmasının ise düşük frekans bandındaki enerjinin azalarak plosive kaynaklı gürültülerin yüksek geçirgen filtreye etkin biçimde giderildiğini göstermektedir. Böylece, konuşma sinyalinin doğal tınısı korunurken teknik bozulmalar minimize edilmiştir.

3.2.1.8. Adaptif Ambiyans ve Pürüzsüzleştirme

Önceki aşamalarda temizlenmiş olan sinyal üzerinde, zamansal sürekliliği artırmak ve işlem sonrası oluşabilecek sert geçişleri azaltmak amacıyla adaptif birleştirme ve pürüzsüzleştirme uygulanmıştır. Bu aşamada sinyal, konuşma ve sessizlik bölgelerine göre değişen oranlarda kendi referans kopyasıyla kontrollü biçimde karıştırılmıştır. Konuşma ağırlıklı bölümlerde karıştırma oranı düşük tutulmuş; düşük enerjili veya gürültü maskesiyle etiketlenmiş bölgelerde ise daha yüksek oranda uygulanmıştır. Böylece dışarıdan yeni bir ambiyans eklenmeksizin, sinyalin kendi spektral sürekliliği korunarak dolgunluk benzeri bir etki elde edilmiştir. Karıştırma işleminin ardından düşük geçiren (low-pass) yumuşatma filtresi uygulanarak ani genlik değişimleri ve üst frekanslardaki sertlik azaltılmıştır. Bu işlem, özellikle önceki aşamalarda uygulanan adaptif kazanç ve gürültü azaltma adımlarının oluşturabileceği keskin spektral sınırları yumuşatmayı hedeflemiştir. Elde edilen çıktı (y_final_smooth), zamansal ve spektral açıdan daha dengeli bir yapı göstermiştir.



Şekil 8. Adaptif ambiyans ve pürüzsüzleştirme karşılaştırması

Karmaşık gürültü ve bozulma türlerinin bir arada bulunduğu ham kayıtlarda, birden fazla filtreleme ve gürültü azaltma yönteminin bir arada kullanıldığı çoklu filtreleme ve kombinasyon yaklaşımları kullanılmaktadır. Karmaşık gürültü koşullarında birden fazla filtreleme yaklaşımının birlikte kullanılmasının sinyal-gürültü oranı (SNR) iyileştirmesine katkı sağladığı literatürde belirtilmektedir (Liu, 2022; Ramli vd., 2017). Bu bağlamda, çalışmada uygulanan kontrollü harmanlama ve yumuşatma adımı, önceki gürültü azaltma ve filtreleme işlemlerini tamamlayıcı bir dengeleme aşaması olarak kurgulanmıştır.

3.2.1.9. Metrikler ve A/B Karşılaştırma

Bu aşamada, işlem zincirinin farklı adımlarında elde edilen ses dosyalarının örnekleme oranı (sample rate, SR) bakımından tutarlı hâle getirilmesi ve sürümler arasındaki farkların nesnel biçimde ölçülmesi hedeflenmiştir. Kod, her dosyanın SR değerini analiz ederek örnekleme oranı farklılıklarına bağlı olası hız veya perde sapması riskini belirlemiş ve referans SR değerine güvenli yeniden örnekleme (resampling) uygulanarak, aynı zaman çözünürlüğü ve örnekleme yapısı üzerinden işlem gerçekleştirilmiştir.

Her işlem adımı için elde edilen sinyal ile referans olan filtrelenmiş sürüm arasındaki fark, karşılaştırmalı bir SNR metriği kullanılarak dB cinsinden hesaplanmıştır. SNR metriği, her sürümün referans dosyadan ne ölçüde ayrıştığını nicel olarak ifade eden göreceli bir fark metriğini temsil etmektedir. Bu yaklaşım, işlem zincirinin her adımında sinyal üzerinde yapılan müdahalenin büyüklüğünü karşılaştırmalı olarak değerlendirmeye olanak tanımaktadır. Örnek görüşme kayıtlarına ilişkin sonuçlarda, “Denoised” sürüm “Final” sürüm arasındaki SNR değerlerinde değişim gözlenmiştir. Bu durum, sonraki aşamalarda bilinçli olarak sinyalin gürültüden arındırıldığını ve spektral şekillendirme ile dinamik düzenleme aşamalarından geçirildiğini göstermektedir. SNR değerindeki değişim referans sürüme kıyasla yapılan işlemlerin kümülatif etkisini yansıtmaktadır. Bu bütünleşik yaklaşım, ses üzerindeki toplam etkiyi ortaya koymakla birlikte, kullanıcıya referans ve nihai sürümü dinleme olanağı sunularak nesnel metriklerle elde edilen bulguların işitsel değerlendirme ile birlikte yorumlanması sağlanmıştır.

4. YAPAY ZEKÂ (AI) DESTEKLİ TÜRKÇE TRANSKRİPSİYON TASARIMI

Transkripsiyon aşamasında, dijital sinyal işleme (DSP) teknikleri ile derin öğrenme tabanlı kaynak ayrıştırma yöntemlerini birleştiren hibrit bir yaklaşım benimsenmiş, sinyaldeki gürültü ve teknik bozulmaları azaltarak model için daha temiz bir ön işleme çıktısı sağlamıştır. Temizlenmiş sinyal, Demucs (htdemucs) derin öğrenme tabanlı kaynak ayrıştırma modeli ile işlenmiştir. Model, konuşma ve arka plan bileşenlerini zaman ve frekans alanında eşzamanlı olarak analiz ederek insan sesi bileşenini diğer kaynaklardan ayırtmıştır. Dolayısıyla konuşma bileşeni arka plana göre daha belirgin hâle getirilmiştir. DSP işlemleri ile derin öğrenme tabanlı kaynak ayrıştırma modelin ile oluşturulan tamamlayıcı yapı, transkripsiyon öncesinde vokal öncelikli ve daha dengeli bir kayıt elde edilmesini sağlamaktadır.

Kaynak ayrıştırma sonrasında, uzun süreli Türkçe ses kayıtlarının otomatik transkripsiyonu gerçekleştirilmiştir. Ses dosyası FFmpeg aracılığıyla yönetilebilir segmentlere ayrılarak her segment WhisperModel (Faster-Whisper) GPU destekli modeller kullanılarak ayrı ayrı çözümlenmiştir. Segment çıktıları birleştirilerek hem zaman kodlu hem de tam metin formatında (.txt ve .docx) dışa aktarılmıştır. Bu segmentasyon yaklaşımı ile uzun görüşme kayıtlarının belleği aşmadan ve sistem kaynaklarını daha verimli kullanarak işlenmesi sağlanmıştır.

Sistem performansının doğruluk düzeyini değerlendirmek amacıyla otomatik transkripsiyon çıktıları, araştırmacı tarafından oluşturulan referans metinlerle karşılaştırılmıştır. Değerlendirmede yaygın olarak kullanılan iki temel metrik olan Kelime Hata Oranı (Word Error Rate, WER) ve Karakter Hata Oranı (Character Error Rate, CER) hesaplanmıştır (Fox vd., 2021; Poulos & Valle, 2021). Hesaplamalar jiver kütüphanesi kullanılarak gerçekleştirilmiş ve Türkçe karakterlere yönelik normalizasyon ve biçimsel temizlik adımları uygulanmıştır. Bu doğrultuda elde edilen WER ve CER değerleri, sistemin doğruluğunu nicel olarak değerlendirilmesini sağlamıştır. Bu yapı, sinyal işleme, kaynak ayrıştırma ve otomatik konuşma tanıma aşamalarını tek bir işlem zincirinde bir araya getiren entegre bir Türkçe transkripsiyon tasarımı sunmaktadır.

5. SONUÇ

Bu çalışma, nitel araştırmalarda ses kaydının, veri kalitesini ve analiz güvenilirliğini doğrudan etkileyen metodolojik bir değişken olduğunu göstermektedir. Alan araştırmalarında karşılaşılan çevresel gürültüler, klik, uğultu ve plosive gibi teknik bozulmalar verinin anlaşılabilirliğini, transkripsiyon doğruluğunu ve çözümleme sürecinin tutarlılığını önemli ölçüde etkilemektedir. Bu doğrultuda araştırmacıların kayıt öncesi süreçte mikrofon seçimi, akustik düzenleme ve ortam kontrolü gibi aşamaları sistematik biçimde ele almaları gerekmektedir.

Çalışma kapsamında geliştirilen Python tabanlı hibrit işlem hattı, geleneksel dijital sinyal işleme (DSP) tekniklerini, Demucs destekli derin öğrenme tabanlı kaynak ayrıştırma ve Whisper temelli otomatik transkripsiyon sistemiyle bütünleştirmektedir. İşlem hattı sırasıyla ön işleme, adaptif gürültü azaltma, uğultu ve klik/plosive temizliği, konuşmayı ön plana çıkarma, çok segmentli transkripsiyon ve doğruluk analizi aşamalarından oluşmaktadır. Bu bütünleşik yapı sayesinde konuşma verileri, teknik dilsel açıdan daha temiz ve analiz edilebilir bir metne dönüştürülebilmektedir.

Doğrulama senaryolarında sistem, %79,66-%95 kelime doğruluğu ve %88,58-%97 karakter doğruluğu aralığında performans sergilemiş; genel WER değeri 0.203 ve CER değeri 0.114 olarak hesaplanmıştır. Bu değerlerin yorumlanması açısından belirtmek gerekir ki 0.203 düzeyinde bir WER, 10 kelimeden yaklaşık 2'sinin hatalı tanındığı anlamına gelmektedir. Ancak bu oranın büyük bölümü noktalama, Türkçeye özgü ek yapıları ve agresif normalizasyon farklılıklarından kaynaklanmaktadır. Karakter düzeyinde ise 0.114'lük CER değeri, üretilen metnin yaklaşık %88,6'sının karakter bazında referans transkripsiyonla örtüştüğüne işaret etmektedir. Nitel araştırma pratiği açısından değerlendirildiğinde bu performans, araştırmacının otomatik çıktıyı hızla gözden geçirerek kullanılabilir hâle getirebildiği, yani tam el transkripsiyonuna kıyasla iş yükünü belirgin biçimde azaltan bir doğruluk eşiğine karşılık gelmektedir.

Sonuç olarak çalışma, nitel araştırmalarda ses verisinin işlenmesi ve metne dönüştürülmesi sürecine teknik ve metodolojik bir bütünlük kazandırmaktadır. Geliştirilen sistem, saha koşullarında elde edilen ham kayıtların temizlenmesini, yapay zekâ destekli transkripsiyon süreçleriyle bütünleştirmekte ve yeniden üretilebilir bir işlem hattı sunmaktadır. Adaptif gürültü azaltma, ambiyans dengelemesi ve doğrulama metrikleriyle desteklenen bu yapı, manuel müdahaleyi azaltarak veri işleme sürecini daha tutarlı ve izlenebilir hâle getirmektedir. Bu yönüyle önerilen model, sosyal bilimlerde dijital araçların araştırma pratiğine entegrasyonuna yönelik uygulanabilir bir çerçeve sunmaktadır. Ayrıca yukarıda belirtilen değerler, kontrollü akustik koşullarda oluşturulan kayıtlar üzerinden elde edildiğinden, gerçek saha görüşmelerine doğrudan genellenmeden önce farklı konuşmacı profilleri, lehçe/ağız çeşitliliği ve üst üste binen konuşmaları içeren kayıtlar gibi değişkenler dikkate alınarak işlem hatlarının çalışılması önerilmektedir.

KAYNAKÇA

- Afifah, K., Arijal, M., Retdian, N., & Shima, T. (2018). Experiment on Hum Noise Suppression using N-path Notch Filter. *2018 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 519-522. <https://doi.org/10.1109/ISPACS.2018.8923266>
- Aljumaily, R., & Abdulmohsin, H. (2024). Review: Noise Reduction Techniques for Enhancing Speech. *Iraqi Journal of Science*, 5798-5818. <https://doi.org/10.24996/ijs.2024.65.10.38>
- Bhattacharya, P. C., & Tangsangiumvisai, N. (2017). An octave-scale multiband spectral subtraction noise reduction method for speech enhancement. *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*, 346-350. <https://doi.org/10.1109/SIPROCESS.2017.8124562>
- Bottalico, P. et. al. (2020). Reproducibility of Voice Parameters: The Effect of Room Acoustics and Microphones. *Journal of voice: official journal of the Voice Foundation*, 34(3), 320-334. <https://doi.org/10.1016/j.jvoice.2018.10.016>
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2018). Eğitimde bilimsel araştırma yöntemleri. Ankara: Pegem Akademi Yayınları.
- Corbett, I. (2014). *Mic It!: Microphones, Microphone Techniques, and Their Impact on the Final Mix*. CRC Press.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). SAGE Publications.
- Cui, P., Zhang, J., & Li, T. T. (2021). Research on Acoustic Environment in the Building of Nursing Homes Based on Sound Preference of the Elderly People: A Case Study in Harbin, China. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.707457>
- Demir, Nurettin (2013). *Ankara Örneğinde Ağızların Belgelenmesi İnceleme-Metinler Sözlük*, TDK Yay., Ankara.
- Eargle, J. (2012). *The Microphone Book: From mono to stereo to surround - a guide to microphone design and application* (2. bs). Routledge. <https://doi.org/10.4324/9780080473468>
- Ekici, M. (2019). *Halk Bilgisi (Folklor) Derleme ve İnceleme Yöntemleri*, Geleneksel Yay., Ankara.

- Eldwaik, O., & F. Li, F. (2018). Mitigating Wind Induced Noise in Outdoor Microphone Signals Using a Singular Spectral Subspace Method. *Technologies*, 6(1), 19. <https://doi.org/10.3390/technologies6010019>
- Erdal, K. Y. (2023). Dijital İçerik (Podcast) Üretiminde Ses ve Müzik Teknolojileri. İçinde M. Kaptan (Ed.), *Güzel Sanatlar Üzerine Güncel Konular ve Araştırmalar*. Özgür Yayınları. <https://doi.org/10.58830/ozgur.pub291>
- Fox, C. B., Israelsen-Augenstein, M., Jones, S., & Gillam, S. L. (2021). An Evaluation of Expedited Transcription Methods for School-Age Children's Narrative Language: Automatic Speech Recognition and Real-Time Transcription. *Journal of Speech, Language, and Hearing Research*, 64(9), 3533-3548. https://doi.org/10.1044/2021_JSLHR-21-00096
- Griffin, A., Zorilă, T.-C., & Stylianou, Y. (2015). Improved face-to-face communication using noise reduction and speech intelligibility enhancement. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5103-5107. <https://doi.org/10.1109/ICASSP.2015.7178943>
- Horvat, E. et. al. (2013). The beginner's guide to doing qualitative research: How to get into the field, collect data, and write up your project. Teachers College Press.
- Karasar, N. (2011). Bilimsel araştırma yöntemi. Ankara: Nobel Yayın Dağıtım.
- Konuma, Y., & Asakura, T. (2023). Effects of microphone mounting location and gender on accuracy in speech recognition using a throat microphone. *JASA Express Letters*, 3(9), 095203. <https://doi.org/10.1121/10.0020988>
- Krueger, R. A. (2014). *Focus groups: A practical guide for applied research*. Sage publications. https://books.google.com/books?hl=tr&lr=&id=8wASBAAQBAJ&oi=fnd&pg=PP1&dq=focus+groups+a+practical+guide&ots=XgfNCB7NmX&sig=54PJ7gSe9xd4_9DJTTRVdgdUGUY
- Liu, R. (2022). Speech noise reduction system based on combined filter. *2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*, 611-616. <https://doi.org/10.1109/TOCS56154.2022.10016136>
- MacCallum, J. K., Olszewski, A. E., Zhang, Y., & Jiang, J. J. (2011). Effects of Low-pass Filtering on Acoustic Analysis of Voice. *Journal of voice : official journal of the Voice Foundation*, 25(1), 15-20. <https://doi.org/10.1016/j.jvoice.2009.08.004>
- Magsi, H., Sodhro, A. H., Chachar, F. A., & Abro, S. A. K. (2018). Analysis of signal noise reduction by using filters. *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 1-6. <https://doi.org/10.1109/ICOMET.2018.8346412>
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.
- Nian, Z., Tu, Y.-H., Du, J., & Lee, C.-H. (2021). A Progressive Learning Approach to Adaptive Noise and Speech Estimation for Speech Enhancement and Noisy Speech Recognition. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6913-6917. <https://doi.org/10.1109/ICASSP39728.2021.9413395>
- Öçalan, M. (2006). Ağız Araştırmalarında Bilişim Teknolojilerinin Kullanılması ve Ağız Tezleri İçin Yenilik Önerileri. *Erciyes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 1(20), 169-189.
- Özkurt, C. (2025). Investigation of the Effectiveness of Audio Processing and Filtering Strategies in Noisy Environments on Speech Recognition Performance. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 8(1), 222-247. <https://doi.org/10.47495/okufbed.1457532>
- Özyıldırım, M. C. (2010). *Kapalı Mekânların Akustik Tasarımında Ses Güçlendirme Sistemlerinin Uygulanma Koşulları*. <http://hdl.handle.net/11527/8238>

- Poulos, J., & Valle, R. (2021). Character-based handwritten text transcription with attention networks. *Neural Computing and Applications*, 33(16), 10563-10573. <https://doi.org/10.1007/s00521-021-05813-1>
- Ramli, R. M., Noor, A. O. A., & Abdul Samad, S. (2017). Noise cancellation using selectable adaptive algorithm for speech in variable noise environment. *International Journal of Speech Technology*, 20(3), 535-542. <https://doi.org/10.1007/s10772-017-9425-1>
- Retdian, N., & Afifah, K. (2020). Hum Noise Suppression Using an On-chip N-path Notch Filter. *2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 1-4. <https://doi.org/10.1109/ICECS49266.2020.9294831>
- Rogers, D. et. al. (2020). The City Under COVID-19: Podcasting as Digital Methodology. *Tijdschrift Voor Economische En Sociale Geografie*, 111(3), 434-450. <https://doi.org/10.1111/tesg.12426>
- Roy, S. K., & Paliwal, K. K. (2021). A noise PSD estimation algorithm using derivative-based high-pass filter in non-stationary noise conditions. *EURASIP Journal on Audio, Speech, and Music Processing*, 2021(1), 32. <https://doi.org/10.1186/s13636-021-00220-9>
- Segerstrom, E., Vanfretti, L., Mishra, C., Xu, X., Jones, K. D., & Gardner, R. M. (2022). Using Spectral Flatness to Detect and Label Power System Oscillations in the Presence of Intermittent Broadband Noise. *2022 IEEE Power & Energy Society General Meeting (PESGM)*, 1-5. <https://doi.org/10.1109/PESGM48719.2022.9917166>
- Seidman, I. (2006). Interviewing as qualitative research: a guide for researchers in education and the social sciences (3. ed.). Teachers College Press.
- Taylor, S. J., Bogdan, R. & DeVault, M. J. (2016). Introduction to qualitative research methods: A guide book and resource. 4. Ed. John Wiley & Sons.
- Thompson, C. B., & Panacek, E. A. (1998). Basics of research (part 13): Qualitative research—An example. *Air Medical Journal*, 17(3), 121-124. [https://doi.org/10.1016/S1067-991X\(98\)90110-9](https://doi.org/10.1016/S1067-991X(98)90110-9)
- Titze, I. R., & Winholtz, W. S. (1993). Effect of microphone type and placement on voice perturbation measurements. *Journal of Speech and Hearing Research*, 36(6), 1177-1190. <https://doi.org/10.1044/jshr.3606.1177>
- Torresin, S. et. al. (2020). Acoustics for Supportive and Healthy Buildings: Emerging Themes on Indoor Soundscape Research. *Sustainability*, 12(15), 6054. <https://doi.org/10.3390/su12156054>
- Tracy, S. J. (2013). Qualitative research methods: Collecting evidence, crafting analysis, communicating impact. Wiley-Blackwell.
- Vergili, S. (2017). Kayıt Edilmiş İnsan Sesinde Konuşma Anlaşılabilirliğinin Değerlendirilmesi ve Anlaşılabilirliği Etkileyen Faktörler. *Journal of International Social Research*, 10(50), 1021-1026. <https://doi.org/10.17719/jisr.2017.1731>
- Wang, W., & Yu, G. (2025). A-weighted Short-Time Fourier Transform: Enhanced Acoustic Signal Analysis of Rotating Machinery. İçinde Z. Wang, K. Zhang, K. Feng, Y. Xu, & W. Yang (Ed.), *Proceedings of the TEPEN International Workshop on Fault Diagnostic and Prognostic* (ss. 282-292). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-73407-6_27
- Yin, R. K. (2011). Qualitative research from start to finish. The Guilford Press.
- Yıldırım, A., & Şimşek, H. (2018). *Sosyal bilimlerde nitel araştırma yöntemleri*. Seçkin Yayıncılık.

Araştırma Makalesi / Research Article

Ses Kaydı Yapılan Görüşmelerde Arka Plan Gürültüsü Azaltımı: Kayıt ve Kayıt Sonrası İşleme Stratejileri

Background Noise Reduction In Audio-Recorded Interviews: Recording and Post-Processing Strategies

Mustafa DEMİR , Kadri Yılmaz ERDAL  & İnan GÜMÜŞ 

DOI : [10.63556/ankad.v10i3.471](https://doi.org/10.63556/ankad.v10i3.471)

Geliş/Received: 10/03/2026

Kabul/Accepted: 18/05/2026

EXTENDED ABSTRACT

Introduction

Audio recording quality is a methodologically consequential variable in qualitative research, yet it has received limited systematic attention in the literature. Interviews conducted in field settings are frequently compromised by environmental noise wind, traffic, crowd sounds, and public announcement systems as well as by room-specific acoustic conditions such as reverberation, resonance, and inadequate microphone placement. Speaker-specific factors, including low vocal volume, hearing impairment, speech disfluencies, and reluctance toward recording devices, further reduce intelligibility. These challenges collectively threaten the reliability, transcription accuracy, and analytical consistency of qualitative data. Despite growing recognition of these problems, most methodological guidance remains at the level of general recommendations such as choosing a quiet venue or testing equipment beforehand without offering technically reproducible or computationally grounded solutions.

Commercial tools like iZotope RX can reduce noise well, but they use proprietary, non-transparent workflows. This limits reproducibility and makes it hard to adapt them to specific conditions. In this context, the present study proposes a hybrid, Python-based pipeline for audio enhancement and automatic transcription. This system is designed for qualitative interview recordings. It integrates classical DSP techniques with deep learning for source separation and automatic speech recognition (ASR). The pipeline is modular, transparent, and reproducible. As a result, recording quality becomes a methodological dimension that researchers can manage through open computational tools.

Methods

In this study, an integrated recording scenario was designed to reflect the characteristic noise conditions and acoustic challenges of qualitative field research, with the aim of measuring the performance of the developed processing pipeline. In this context, a dataset derived from standard texts read aloud by researchers was used to systematically evaluate the impact of post-recording processing strategies on noise reduction performance and transcription accuracy. This controlled design enabled the systematic evaluation of the proposed improvement chain's impact on noise reduction and transcription accuracy within a framework of complexity similar to real-world field conditions.

The processing pipeline was implemented in Python and made publicly accessible through a Gradio-based web interface hosted on Hugging Face Spaces. The pipeline comprised eight sequential DSP stages. In the first stage, audio files in various formats (WAV, MP3, FLAC, M4A) were loaded using librosa, soundfile, and pydub, and Short-Time Fourier Transform (STFT)-based spectrograms were generated to establish a spectral baseline. In the second stage, frame-level energy density and spectral

flatness were computed jointly to identify noise-dominant regions and construct an adaptive noise mask, which was subsequently smoothed via Gaussian filtering. The third stage applied masked notch filtering to suppress electrical hum and its harmonics, with the fundamental hum frequency automatically detected within the 40–200 Hz range using `scipy.signal.iirnotch`, applied exclusively in noise-labeled regions to preserve the natural character of speech.

The fourth stage performed fully adaptive spectral noise reduction, where per-frame SNR and spectral centroid estimates drove dynamic suppression levels to prevent over-suppression and timbre degradation in low-energy speech segments, with temporal smoothing applied between frames to ensure continuity. The fifth stage applied dialog isolation through adaptive gain shaping and a Presence EQ targeting the 2.8–6.5 kHz range, foregrounding speech-critical frequencies while attenuating low-frequency background energy. The sixth stage addressed transient artifacts clicks and plosive bursts using derivative-based amplitude change detection and zero-phase Butterworth high-pass filtering via `filtfilt`. The seventh stage applied adaptive blending and low-pass smoothing to reduce residual processing artifacts and soften abrupt spectral transitions introduced by earlier stages. Finally, the eighth stage computed a comparative SNR metric across processing versions to quantify the cumulative impact of each intervention.

Following DSP processing, the enhanced signal was passed through the Demucs (`htdemucs`) deep learning source separation model, which analyzed audio in both the time and frequency domains to isolate the vocal component from residual background sources. The separated vocal track was then segmented using FFmpeg and transcribed using the Faster-Whisper model with Turkish language settings. Outputs were exported in both timestamped and plain-text formats (.txt and .docx). Transcription accuracy was evaluated against manually prepared reference transcriptions using Word Error Rate (WER) and Character Error Rate (CER), computed via the `jiwer` library with Turkish character normalization applied.

Results

The Python-based hybrid processing pipeline developed as part of this study integrates traditional digital signal processing (DSP) techniques with Demucs-supported deep learning-based source separation and a Whisper-based automatic transcription system. The processing pipeline consists of the following stages: pre-processing, adaptive noise reduction, hum and click/plosive removal, speech enhancement, multi-segment transcription, and accuracy analysis. Thanks to this integrated structure, speech data has been converted into text that is technically cleaner and more analyzable.

In validation scenarios, the system demonstrated performance within the ranges of 79.66%–95% word accuracy and 88.58%–97% character accuracy; the overall WER was 0.203, and the CER was 0.114. In interpreting these values, it should be noted that a WER of 0.203 implies that approximately 2 out of every 10 words are misrecognized. However, the majority of this rate stems from punctuation, Turkish-specific suffix structures, and differences in aggressive normalization. At the character level, the CER value of 0.114 indicates that approximately 88.6% of the generated text matches the reference transcription on a character-by-character basis. From a qualitative research perspective, this performance meets a threshold of accuracy, allowing the researcher to quickly review the automated output and render it usable, thereby significantly reducing the workload compared to a full manual transcription.