

Sağlıklı ve Patolojik Seslerin Yapay Sinir Ağları Kullanarak Sınıflandırılması Classification of Healthy and Pathological Voices Using Artificial Neural Networks

Ramis İLERİ, Fatma LATİFOĞLU, Ayşegül GÜVEN

Erciyes Üniversitesi, Mühendislik Fakültesi, Biyomedikal Mühendisliği Bölümü, Kayseri/TÜRKİYE
e-mail : ramissileri@gmail.com, flatifoglu@erciyes.edu.tr, aguven@erciyes.edu.tr

Özetçe— Konuşma insanlar arasındaki iletişimin temelini oluşturmaktadır. Günlük hayatta meslek, çevre, yaş, cinsiyet vb. nedenlerden dolayı sesi meydana getiren mekanizmalarda fonksiyon kayıpları meydana gelmektedir. Bu çalışmada 57 sağlıklı ve 150 patolojik ses sinyallerinin (Hiperkinetik Disfoni, Hipokinetik Disfoni, Reflü Larenjiti hastaları) ortaya konulan öznelikler kullanılarak Yapay Sinir Ağları (YSA) ile sınıflandırılması ele alınmaktadır. Oluşturulan YSA modeline ses sinyallerinden elde edilen özellikler giriş olarak uygulanmıştır. Bu giriş verilerine karşılık yapay sinir ağının çıkış bilgisi, hasta veya sağlıklı iki grup olarak belirlenmiştir. Tasarlanan YSA hasta ve sağlıklı olarak sınıflandırma işleminin en az hata ile gerçekleştirilmesi için iki gizli katman ve bir çıkış katmanı kullanılmıştır. Çalışma sonucunda hasta ve sağlıklı gruplar arasında sınıflandırma başarıları %90.47 olarak elde edilmiştir.

Anahtar Kelimeler — Yapay Sinir Ağları, Patolojik Ses, Sınıflandırma, Sinyal İşleme

Abstract— Speech is the basis of communication between people. In daily life, function loss occurs in mechanisms that create voice due to reasons such as occupation, environment, age and gender. In this study, 57 healthy and 150 pathological voice data (from Hyperkinetic Dysphonia, Hypokinetic Dysphonia, Reflux Laryngitis) was classified using proposed features and Artificial Neural Networks (ANN). The data obtained from voice data is given as input to ANN model. Depending on these input data, the output information of the artificial neural network is determined as patient or healthy. In order to classify the patient and healthy group, two hidden layers and an output layer were used as the artificial neural network model with the least error. As a result of the study, the accuracy of classification between patients and healthy groups was 90.47 %.

Keywords — Artificial Neural Networks (ANN), Pathological, Voice, Classification, Signal Processing

I. GİRİŞ

Sosyal hayatın en temel gereksinimlerinden olan iletişim günlük yaşamımızın büyük bölümünü oluşturmaktadır. Teknolojinin gelişmesiyle sözlü iletişim yerini yazılı iletişime bırakıyor gibi görünse de sözlü iletişim hala insanlar arasındaki etkili iletişim aracıdır. Sözlü iletişim insanların birbirleri ile konuşması ile gerçekleşir. Konuşma anatomik olarak insan vücudundaki bazı mekanizmaların sesi meydana getirmesiyle oluşur. Bu açıdan ele alındığında ses insan yaşamı için önem arz etmektedir. İnsan yaşamı için bu denli önemli olan sesi meydana getiren mekanizmalarda meydana gelebilecek

herhangi bir aksaklık bireyin günlük eylemlerini, hayatını zora sokabilmektedir. Bundan dolayı ses patolojisinde meydana gelen hastalıkların erken teşhisi önem arz etmektedir.

Bu çalışma ile sağlıklı ve patolojik ses verilerinin YSA yöntemleri ile sınıflandırılması amaçlanmaktadır. Literatürde patolojik ve sağlıklı ses verileriyle yapılan çalışmalar incelenecek olursa; normal seslerin patolojik seslerden ayırt edilmesinde Pitch, Formants, Jitter ve Shimmer [1] gibi akustik parametrelerin yeteneği araştırılmıştır. Patolojik ses ile normal sesin sınıflandırılmasında, destek vektör makinesi kullanılarak gerçekleştirilen çalışmalar bulunmaktadır. Sınıflandırma doğruluğunu artıran en iyi özellik setini seçmek için Genetik Algoritma kullanılmıştır. Mel frekanslı hücre katsayısı ve Gauss Karışım Modeli kullanılarak konuşmanın akustik özelliklerine odaklanan bir sınıflandırma tekniği önerilen çalışmalar literatürde yer almaktadır [2]. Konuşmaların normal ve patolojik olarak sınıflandırılması için gizli markov modeline dayanan bir yöntem kullanılmıştır. Bu iki sınıfa ait iki gizli markov modeli eğitilmiş ve eğitilmiş modeller verisini sınıflandırmak için kullanılmıştır [3]. Literatürdeki diğer çalışmalar incelendiğinde F_0 olarak adlandırılan glottisin açılıp kapanma sıklığını belirten temel frekansını belirleme yöntemi.



Şekil 1- Glottis'in Yapısı

Yapılan çalışmaların genelinde sinyallerin belirli zaman aralıklarında otokorelasyonları kullanarak temel frekansını belirlenmesi amaçlanmıştır. Laura ve arkadaşlarının gerçekleştirdiği bir çalışmada glottisin titreşim frekansını tahmin etmek için yeni bir yöntem önermiş ve elde ettikleri temel frekansları patolojik ve sağlıklı sesleri sınıflandırmada kullanmıştır. Bu çalışmada sınıflandırma sonucunda % 76.43 doğruluk elde edilmiştir [4].

Literatürdeki diğer bir çalışmada otokorelasyon ve ortalama genlik fark fonksiyonu kullanarak ses sinyallerinden özellikler elde edilmiş ve bu özellikler sınıflandırmada kullanılmıştır. Bu çalışma sonucunda % 79.02 doğruluk elde edilmiştir [5]. Literatürde glottis temel frekansını belirlemek için önerilen

farklı yöntemler de mevcuttur. Bu yöntemlerden bir tanesinde Gaston Schlotthauer ve arkadaşları patolojik ve sağlıklı sesleri sınıflandırmak için AKA (Ampirik Kip Ayrışımı) yöntemini kullanmışlardır [6]. Gerçekleştirilen çalışmada, patolojik ve sağlıklı bireylere ait ses sinyallerinden zaman ve frekans ekseninden elde edilen özellikler kullanılarak YSA ile sınıflandırılmıştır.

II. YÖNTEM

Çalışma gerçekleştirilmesi sırasında izlenen adımlar Şekil.2.'de belirtilmiştir. Bu adımlar ilerleyen bölümlerde detaylandırılacaktır.



Şekil 2. Çalışmada İzlenen Adımlar

A. Veriseti

Çalışmada kullanılan ses sinyalleri 'PHYSIONET' isimli veritabanında bulunan VOICED (VOice I Car fEDerico II) verisetinden alınmıştır [7]. Bu veritabanı, 150 patolojik ve 57 sağlıklı bireyden alınan toplamda 207 ses kaydı içermektedir. Veritabanındaki ses sinyalleri, herhangi bir kesinti olmadan, beş saniye boyunca sesli harflerin sesli seslendirilmesi sırasında kaydedilmesiyle elde edilmiştir. Veritabanında cinsiyet, yaş, patoloji, yaşam tarzı alışkanlıkları (örneğin sigara, alkol ve kahve tüketimi), mesleki durum ve iki spesifik tıbbi anketin sonuçları gibi bilgileri (Sesli Handikap Endeksi (VHI) ve Reflü Belirti Endeksi (RSI) bilgileri yer almaktadır [1].



Şekil 3.a. Sağlıklı Ses Sinyali
Şekil 3.b. Patolojik Ses Sinyali

Tüm sinyaller 8000 Hz örnekleme frekansı ile örneklenmiştir. Çalışmaya katılan bireyler 18 ile 70 yaşları arasında yetişkinlerden oluşmaktadır. Bu çalışmada 18 yaşın altındaki ve 70 yaşın üzerindeki kişiler ve üst solunum yolu enfeksiyonları veya nörolojik rahatsızlıklar gibi hastalıkları olan kişiler hariç tutulmuştur. Veriseti hakkında daha fazla bilgi ref.[1].

B. Patolojik Ses Kayıtları

Bu çalışmada kullanılan patolojik ses sinyalleri Hiperkinetik Disfoni, Hipokinetic Disfoni, Reflü Larenjiti hastalıklarından kaydedilmiştir.

C. Sinyallerden Elde Edilen Özellikler

Oluşturulan YSA modeline giriş olarak uygulanan özellikler ses sinyallerinden elde edilen varyans, enerji, maksimum güç, entropi, çarpıklık, ses sinyallerinin AKA elde edilen ikinci altband sinyalinin maksimum gücü, maksimum frekansı ve enerjisidir. Bu özelliklere ek olarak verisetinde katılımcılara yapılan anket sonucunda elde edilen VHI ve RSI bilgileri de giriş olarak uygulanmıştır. Veriler min-max normalizasyonu ile normalize edilmiştir ayrıca veriler YSA modeline giriş olarak uygulanmadan önce verilere Temel Bileşenler Analizi (TBA) uygulanmıştır. 6 önemli bileşen kullanılmıştır. Her bir bileşenin sırasıyla %33.6, %20.5, %18.3, %13.1, %2.9 ve %2.6 varyansa sahiptir.

C.1. Ampirik Kip Ayrışımı

AKA metodu ile herhangi bir sinyal Esas Mod Fonksiyonları (EMF) adı verilen bileşenlere ayrıştırılır. Ayrıştırılan sinyal, EMF bileşenlerinin ve artık sinyalin toplamı olarak gösterilmektedir (Eşitlik-1) [3]. AKA algoritması aşağıdaki gibi özetlenebilir;

$$x(t) = \sum_{i=1}^k EMF_i + r(t) \quad (1)$$

- İlk adım maksimum ve minimum noktalarının veya sıfır geçiş noktalarının belirlenmesidir.
- İkinci adımda; maksimumların kübik interpolasyon ile birleştirilmesi sonucu sinyalin üst zarfı elde edilir ve minimumlar içinde alt zarf elde edilmektedir.
- Üçüncü adımda; bütün zaman dilimi için alt ve üst zarfların ortalaması elde edilir. Zaman eksenine işaretine $X(t)$ ve ortalama işaretine $m1$, sinyalin üst zarfı $zmax$ ve alt zarfı $zmin$ olarak ifade edilirse aşağıdaki eşitlikle elde edilir;
$$m1(t) = (zmax(t) + zmin(t)) / 2 \quad (2)$$
- Dördüncü adımda; ortalama sinyal asıl sinyalden çıkarılarak detay sinyal elde edilir. Detay sinyal $d(t)$ olarak aşağıdaki eşitlikle ifade edilir;
$$[d(t) = x(t) - m(t)] \quad (3)$$
- Son adımda; detay işaretini sanki yeni orijinal sinyalmiş gibi ele alarak tanımlanan işlem basamaklarını durdurma kriterleri sağlanana kadar tekrar edilir.

C.2. Güç Spektral Yoğunluğu

Bir sinyalin frekans spektrumunun incelenmesinde Fourier dönüşümü kullanılmaktadır. Ancak Fourier dönüşümü gürültüye karşı hassas olma dezavantajına sahiptir. Bu nedenle Güç Spektral Yoğunluğu (GSY) yöntemi daha fazla tercih edilmektedir. GSY sinyalin otokorelasyonunun Fourier Dönüşümünün alınması olarak ifade edilir (Eşitlik-4) [8].

$$GSY = \int_{-\infty}^{\infty} |X(t)|^2 e^{-j\omega t} dt \quad (4)$$

C.3. Entropi

Entropi sinyaldeki düzensizliğin bir ölçüsüdür. Bir sinyalin entropisi ise aşağıdaki gibi hesaplanmaktadır. (Eşitlik-5)

$$H(X) = -\sum_{i=1}^M P(X=x_i) \cdot \log_2 P(X=x_i) \quad (5)$$

C.4. Enerji

Sinyalin enerjisi aşağıdaki gibi hesaplanmaktadır: (Eşitlik-6)

$$E = \int_{-\infty}^{\infty} |X(t)|^2 dt \quad (6)$$

C.5. Varyans

Varyans bir serideki değerlerin ortalamaya göre değişimi hakkında bilgi vermektedir. Standart sapmanın karesi alınarak hesaplanır. (Eşitlik-7)

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} \quad (7)$$

D. Temel Bileşenler Analizi

TBA, çok değişkenli bir veriseti içerisindeki bilgiyi daha az değişkenle ve minimum bilgi kaybıyla açıklamanın bir matematiksel tekniğidir. Başka bir tanımla TBA, çok sayıda birbiri ile ilişkili değişkenler içeren verisetinin boyutunu, veriseti içerisindeki veriyi koruyarak daha küçük boyuta indirgenmesini sağlayan bir dönüşüm tekniğidir. TBA, büyük boyutlu verisetlerindeki boyutsallığı azaltır. Teknik, boyut küçültme işlemi veriseti içerisindeki değişken sayısını azaltmayı hedefler. Dönüşüm sonrasında elde edilen değişkenler ilk değişkenlerin temel bileşenleri olarak adlandırılır. İlk temel bileşen olarak varyans değeri en büyük olan seçilir ve diğer temel bileşenler varyans değerleri azalacak şekilde sıralanarak en önemli bileşen seçilir[9].

E. Yapay Sinir Ağları

YSA, beyin gibi biyolojik sinir sistemlerinin bilgi işleme yöntemlerinden ilham alan matematiksel modellerdir. Birbirine bağlı olarak çalışan nöronlardan oluşan YSA mimarisi veri sınıflandırması gibi özel uygulamalar için yapılandırılmışlardır [10]. Bir yapay sinir ağı temel olarak üç katmandan oluşur. Bu katmanlar; girdi katmanı, ara katmanlar, çıktı katmanı olarak adlandırılır.

Girdi katmanından ağa verilen bilgiler, ara katmanlarda ağırlık değerleri kullanılarak çıktı fonksiyonuna dönüştürülür ve çıktı katmanına gönderilir. Girdiler için ağırlık doğru çıktıları üretebilmesi, ağırlık değerlerinin doğru bir şekilde güncellenmesine bağlıdır. Başlangıç ağırlık değerleri rastgele atanır, sonra her bir örnek ağa gösterilerek ağırlık öğrenme kuralına göre ağırlıklar değiştirilir. Bu işlem ağırlık eğitimi olarak adlandırılır. Ağırlıkların güncellenmesi işlemi ile birlikte doğru çıktılar üretilince test setindeki örnekler ağa gösterilir. Eğer ağ test setindeki örneklere doğru cevaplar verirse ağ eğitilmiş kabul edilmektedir [11].

Gerçekleştirilen çalışmada Çok Katmanlı Algılayıcı (ÇKA) kullanılmıştır. ÇKA'lar, ileri beslemeli, tam bağlantılı, bir girdi katmanı, bir veya birden fazla gizli katman ve bir çıktı katmanından oluşan hiyerarşik yapıya sahip yapay sinir ağı mimarileridir. Giriş katında herhangi bir bilgi işleme yapılmaz. Buradaki işlem elemanı sayısı problemin giriş sayısına bağlıdır. Ara katman sayısı deneme yanılma yoluyla bulunur. Çıkış katmanı eleman sayısı ise uygulanan probleme bağlıdır. Çok

katmanlı algılayıcılar, bilgi girişinin yapıldığı girdi katmanı, bir veya daha fazla gizli (ara) katman ve bir çıktı katmanından oluşmaktadır. Çok katmanlı algılayıcılarda katmanlar arası ileri ve geri yayılım olarak adlandırılan geçişler bulunur. İleri yayılım safhasında, ağırlık çıktısı ve hata değeri hesaplanır. Geri yayılım safhasında ise hesaplanan hata değerinin minimize edilmesi için katmanlar arası bağlantı ağırlık değerleri güncellenir. Yapılan çalışmada giriş katmanı, gizli katman ve çıkış katmanı ile oluşturulan 3 katmanlı ağ yapısı ve geri yayılım öğrenme algoritması kullanılmıştır.

57 sağlıklı ve 150 patolojik ses sinyallerinden elde edilen 10 özellik ÇKA yapısına giriş olarak uygulanmıştır. Ağırlık çıkışında sağlıklı veya patolojik olmak üzere iki sınıf elde edilmiştir. Oluşturulan ağ modelinde iki adet gizli katman kullanılmıştır. Deneysel olarak farklı çıkış sayıları denenmiş ve en iyi sonucun elde edildiği ilk gizli katman için çıkış sayısı 6, ikinci gizli katman için ise 5 olarak belirlenmiştir. Çıkışta 0 (sağlıklı ses) ve 1 (patolojik ses) bilgisi elde edilmiştir. Verisetinin bir kısmı ile yapay sinir ağı eğitilmiş, bir kısmı ile eğitilen yapay sinir ağının başarısını ölçülmüştür. Yapay sinir ağlarının eğitimi ve test edilmesi aşamalarında kullanılacak sinyallerin belirlemek için K-katlamalı çapraz doğrulama yöntemi kullanılmıştır. K-katlamalı çapraz doğrulama, verisetini parçalara ayırma yöntemlerinden biridir. K-katlamalı çapraz doğrulama, veriyi belirlenen bir K sayısına göre eşit parçalara böler, her bir parçanın hem eğitim hem de test için kullanılmasını sağlar, böylelikle dağılım ve parçalanmadan kaynaklanan sapma ve hataları minimuma indirir [12].

Bu çalışmada K değerleri 5, 6, 7, 8, 9 ve 10 olarak seçilmiş ve her bir K değeri için oluşturulan ağ 3 kez çalıştırılmış ve ortalama değerleri alınmıştır. Böylece en yüksek doğruluğun elde edildiği K değeri belirlenmiştir.

III. SONUÇLAR VE TARTIŞMA

İletişimin temeli olan konuşma insanların temel fonksiyonlarından biridir. Günlük hayatta meslek, çevre, yaş, cinsiyet vb. nedenlerden dolayı sesi meydana getiren mekanizmalarda fonksiyon kayıpları meydana gelmektedir. Bu durumdan etkilenen kişi sayısı giderek artmaktadır. Artan hasta sayısına paralel olarak doktorlara düşen iş yükü de artmaktadır. Ses rahatsızlıklarının teşhisini kolaylaştırarak doktorların iş yükünü hafifletmek amacıyla çeşitli yöntemler geliştirebilmek adına yapılan çalışmalar bu çalışmanın önemini göstermektedir. Yapılan çalışmada, yapay sinir ağları kullanılarak sağlıklı ve patolojik seslerin sınıflandırılması amaçlanmıştır. Bu amaç doğrultusunda yapılan çalışmalar iki kısma ayrılmıştır. Birinci kısım, ses sinyallerinden öznelik çıkarımı; ikinci kısım ise yapay sinir ağları ile sınıflandırma işlemidir. Literatürde patolojik sesleri sınıflandırmak için gerçekleştirilmiş çok sayıda çalışma bulunmaktadır [13-15]. Önerilen bu çalışmada VOice I Car fEDerico II isimli veritabanından toplamda 207 kişiye ait ses sinyalleri kullanılmıştır.



Sınıflandırıcı doğruluk değerleri aşağıdaki formülle hesaplanmıştır;

$$\text{Doğruluk} = \frac{DN + DP}{(DP + DP + YN + YP)}$$

DN: Doğru Negatif
DP: Doğru Pozitif
YN: Yanlış Negatif
YP: Yanlış Pozitif

YSA modeline farklı K-katlamalı çapraz doğrulama değerleri verilmiştir. Tablo 1'de görüldüğü gibi sınıflandırıcı performansı en yüksek K=10 katlamalı çapraz doğrulama için elde edilmiş ve %90.47 olarak hesaplanmıştır.

Tablo 1. Farklı K Değerleri için Doğruluk Değerleri

K Değeri	Doğruluk
K=5	%76.19
K=6	%76.47
K=7	%83.33
K=8	%80.76
K=9	%78.26
K=10	%90.47

Tablo.2'de literatürde gerçekleştirilen diğer çalışmalardan elde edilen sınıflandırıcı performansları verilmiştir. Bu çalışmalar patolojik ses sinyallerinin sınıflandırılmasında çeşitli yöntemler önermişlerdir. Önerilen bu çalışmanın literatürdeki çalışmalardan farkı temel sinyal özellikleri kullanarak ses verilerini sınıflandırmaktır.

Tablo 2. Literatür Karşılaştırma

Algoritma	Doğruluk(%)
Önerilen Çalışma	%90.47
F ₀ Tahmini [4]	76.43
PRAAT [5]	70.71
AMDF [16]	68.57
SWIPE [17]	67.86
SHRP [18]	67.86
YIN [19]	72.97
RAPT [20]	71.43
STRAIGHT [21]	67.86

Bu çalışmada, ses sinyallerinin zaman ve frekans eksenindeki temel özellikleri, AKA uygulanarak elde edilen alt band sinyallerin özellikleri ve verisetinden elde edilen özellikler (VHI ve RSI) kullanılarak sınıflandırma işlemi yapılmıştır. Elde edilen sonuçlar literatürde incelenen çalışmalara kıyasla daha yüksektir. Bu çalışma sinyalin zaman ve frekans eksenindeki temel özellikleri (GSY, Enerji vb.) ile patolojik ses sinyallerinin sınıflandırılmasında kullanılan bir ön çalışmadır. Bu bulgulardan yola çıkarak sınıflandırıcı başarısı arttırmak üzere yapılacak olan yeni çalışmalarda ses sinyalleri üzerinde farklı sinyal işleme yöntemleri denenerek çalışmalara devam edilmesi planlanmaktadır.

KAYNAKLAR

- [1] U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, and L. Verde. A new database of healthy and pathological voices. Computers & Electrical Engineering, vol. 68, pp. 310-321, 5 2018
- [2] Ugo C., Giuseppe P., Elio M., Ciro N., Givovanna S., "A new database of healthy and pathological voice" ,Computer and Electrical Engineering 2018,310:321
- [3] Majidnezhad V., Kheidorov I., "A HMM-Based Method for Vocal Fold Pathology Diagnosis", International Journal of Computer Science, Vol. 9, Issue 6, No 2, November 2012.
- [4] Verde L., De Pietro G., Sannino G., "A methodology for voice classification based on the personalized fundamental frequency estimation", Biomed Signal Processing and Control 2018;42:134
- [5] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound" Proceedings of the Institute of Phonetic Sciences, vol. 17, Amsterdam (1993) 97-110
- [6] Schlotthauer G., Torres E., Rufiner E., "Pathological Voice Analysis and Classification Based on Empirical Mode Decomposition", Training School 2009 ,LCSN 5967,364-381
- [7] <https://physionet.org/pn6/voiced/>
- [8] K.L. Wu, P.F. Hsieh, "Empirical mode decomposition for dimensionality reduction of hyperspectral data", IEEE International Geoscience and Remote Sensing Symposium, vol. 2, s. 1241- 1244, 2005.
- [9] <https://www.bilgiinegi.com/temel-bilesen-analizi-principal-component-analysis-nedir.htm>, Erişim Tarihi: Temmuz 2019
- [10] Erkaymaz, H., Özer, M. ve Orak, İ. M., "Detection of Directional Eye Movements Based on The Electrooculogram Signals Through an Artificial Neural Network", Chaos Solitons&Fractals, 56 : 202- 208, 2015.
- [11] Öztemel, E., (2012). Yapay Sinir Ağları. İstanbul: Papatya Yayıncılık Eğitim.
- [12] <http://www.datascience.istanbul/2017/08/29/bir-bakista-k-fold-cross-validation/> (Erişim Tarihi: Mayıs 2019)
- [13] Carvalho,R.T.S ,Cavalcante,C, Cortez,P,C," Wavelet Transform and Artificial Neural Networks Applied to Voice Disorders Identification", 2011 Third World Congress on Nature and Biologically Inspired Computing,371-376
- [14] Salhi,L,Mourad,T, Cherif,A,"Voice Disorders Identification Using Multilayer Neural Network", The International Arab Journal of Information Technology, No. 2, April 2010
- [15] Srinivasan,V,Ramalingam,V, Arulmozhi,V," Artificial Neural Network Based Pathological Voice Classification Using MFCC Features, International Journal of Science, Environment ISSN 2278-3687 (O) and Technology, No 1, 2014, 291 – 302
- [16] M.J. Ross, H.L Shaffer, A. Cohen, R. Freudberg, H.J. Manley, Average magnitude difference function pitch extractor, IEEE Trans. Acoust. Speech Signal Process. 22 (5) (1974).
- [17] A. Camacho, Swipe: a sawtooth waveform inspired pitch estimator for speech and music (Ph.D. thesis), University of Florida, 2007.
- [18] X. Sun, Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio, in: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 1, IEEE, 2002, pp. 1-333.
- [19] A. De Cheveigné, H. Kawahara, Yin, a fundamental frequency estimator for speech and music, J. Acoust. Soc. Am. 111 (4) (2002) 1917-1930.
- [20] D. Talkin, A robust algorithm for pitch tracking (RAPT), in: Speech Coding and Synthesis, 1995.
- [21] H. Kawahara, A. de Cheveigné, H. Banno, T. Takahashi, T. Irino, Nearly defect-free f0 trajectory extraction for expressive speech modifications based on STRAIGHT, Interspeech (2005) 537-540