

# Öğrenme Aktarmanın Çok-Kipli Konuşma Analizi için Kullanılması



KOÇ  
ÜNİVERSİTESİ

Mehmet Ali Tuğtekin Turan , mturan@ku.edu.tr

## Motivasyon

- Akustik Mikrofon: Yüksek kaliteli kayıt fakat gürültülerden kolay etkilenme
- Gırtlak Mikrofonu: Gürültüye karşı gürbüz ancak dar-bantlı konuşma kayıtları
- Gırtlak mikrofonu kayıtlarının kalitesini ve doğallığını arttırmak mümkün mü? <sup>1</sup>



## Öğrenme Aktarma (Transfer Learning)

- Geleneksel öğrenimde veri dağılımları, öznitelik ve çıktı uzayları aynı olmalıdır
- Öğrenme aktarma her görevi sıfırdan öğrenmek yerine önceki görevleri kullanır

## Gürültü Giderici Yağın Otokodlayıcı (GGYO)

- Kaynak  $[K, K^{(p)}]$  ve hedef  $[H, H^{(p)}]$  alanı için karesel hata maliyetini en aza indirgeyecek, ağırlık matrisleri,  $W_K, W_H$ ,

$$\sum_i \left\| [K, K^{(p)}] - W_K [K, K^{(p)}] \right\|^2$$

burada (i) gürültü ile bozulmuş veridir

- GGYO çıktısıyla doğrusal olmayan öznitelikler,  $[U_K, U_K^{(p)}] = \tanh(W_K [K, K^{(p)}])$
- Her katman için ağırlık matrisleri  $W_K, W_H$  yeni öznitelik gösterimleri  $U_K, U_H, U_K^{(p)}, U_H^{(p)}$
- Herhangi bir  $l$  katmanında öğrenilecek öznitelik dönüşüm matrisi,

$$G_l = \left( U_{K,l}^{(p)} (U_{H,l}^{(p)})^T \right) \left( U_{H,l}^{(p)} (U_{H,l}^{(p)})^T + \lambda I \right)^{-1}$$

- Dönüştürülmüş öznitelikler,

$$D_H^{(p)} = [(G_1 U_{H,1}^{(p)}), (G_2 U_{H,2}^{(p)}), \dots, (G_L U_{H,L}^{(p)})]^T$$
$$D_K = [U_{K,1}, U_{K,2}, \dots, U_{K,L}]^T$$

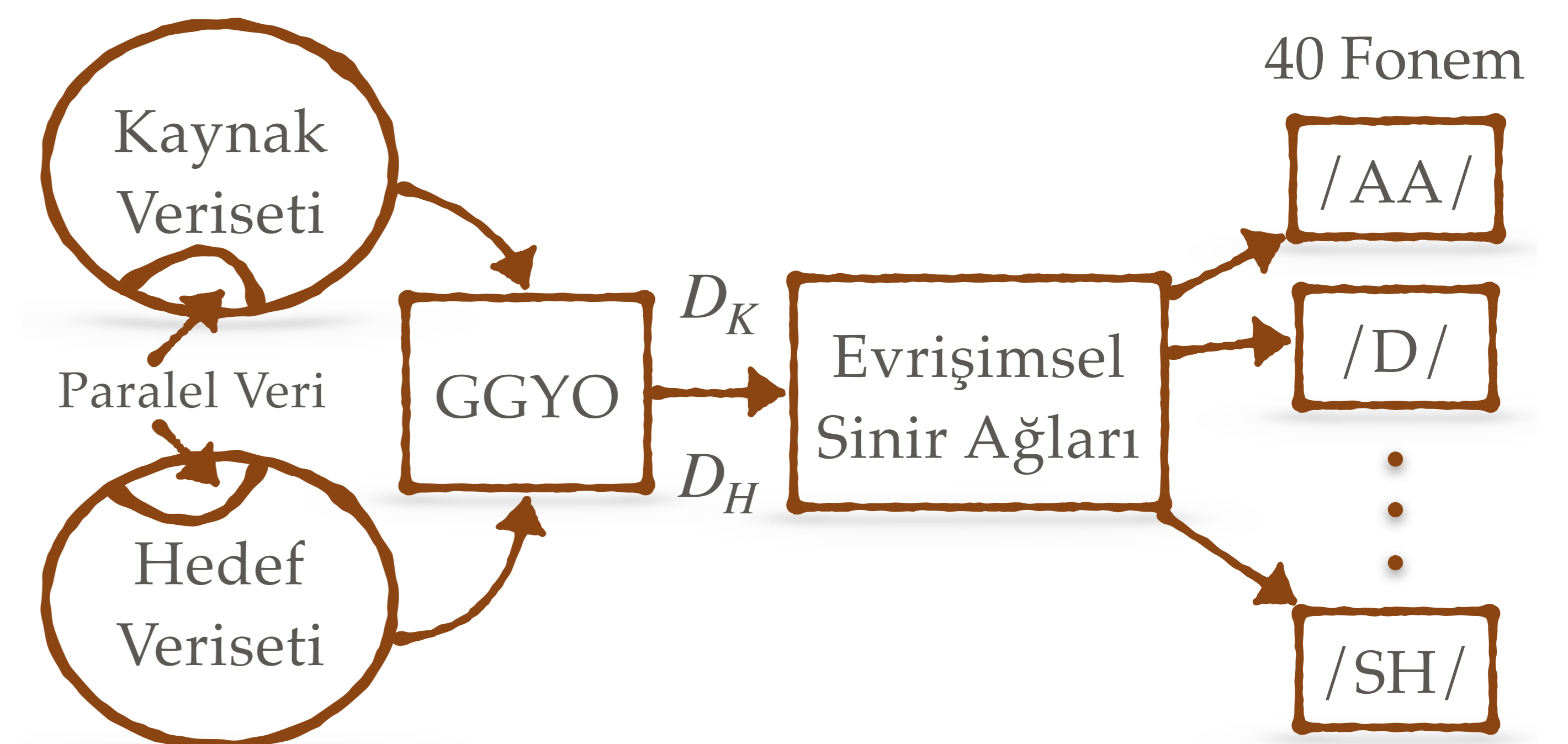
## Veriseti

- Kaynak Veriseti: 120 kullanıcıdan, 2462 cümlelik Türkçe veritabanı, "METUbet" <sup>2</sup>
- Paralel Veriseti: Tek kullanıcıyla eşzamanlı kaydedilmiş, 800 cümleden oluşan akustik ve gırtlak mikrofonu kayıtları

## Öznitelik Dönüşümü

- Süzgeç kümesi (filter-bank) öznitelikleri 40 adet mel bandıyla, içiçe geçen 25 ms'lik çerçeveler üzerinden hesaplandı
- Toplamda 40 farklı foneme ait kaynak ve hedef verisetlerine GGYO uygulandı
- GGYO ile iyileştirilmiş öznitelikler, evrişimsel sinir ağları ile sınıflandırıldı

## Genel Şema



## Fonem Sınıflandırma Sonuçları

Deney	Sınıflandırma Başarımı
Sadece Gırtlak Mikrofonu	% 31
Sadece Akustik Mikrofon	% 78
GGYO İyileştirmesi	% 59

## Referanslar

- 1 Tuğtekin Turan ve Engin Erzin. "Source and Filter Estimation for Throat-Microphone Speech Enhancement" IEEE Transactions on Audio, Speech and Language Processing, 2016
- 2 Tolga Çiloğlu, Mübeccel Demirekler ve diğerleri "Developing New Audio Corpora and Speech Recognition Tools for the Turkish Language" 2002