Edibe YILMAZ

Türkçe ve Sosyal Bilimler Eğitimi Bölümü, Hacettepe Üniversitesi

TRO703: Türkçe Eğitimi Araştırmalarıi

**RAPOR**

**Ciferri, M., Ferrante, M., & Toschi, N. (2024). *Bridging auditory perception and language comprehension through MEG-driven encoding models* (arXiv:2501.03246). arXiv.** [**https://doi.org/10**.**48550/arXiv.2501.03246**](https://doi.org/10.48550/arXiv.2501.03246)

**Konu:** Konuşma temelli işitsel uyaranların işlenmesinde işitsel algı ile dilsel anlamanın nöral mekanizmalarını ayırt etmek amacıyla hesaplamalı nörobilim çerçevesinde MEG tabanlı kodlama modelleriyle beyin aktivitesinin modellenmesi.

**Anahtar Kelimeler:** Manyetik ensefalografi (MEG), kodlama modelleri, işitsel algı, dil anlama, konuşma işleme, nöral temsiller, eğitilmiş dil modelleri

**Araştırmanın Amacı ve Gerekçesi**: Dilin beyinde işlenme mekanizmalarını anlamak, bilişsel nörobilimin merkezinde yer alan bir hedeftir; ancak bu süreçleri inceleyen önceki çalışmalar çoğunlukla düşük zaman çözünürlüklü fMRI verilerine dayandığından, konuşma gibi hızlı değişen uyaranların nöral dinamiklerini yeterince ayrıntılı yakalayamamıştır. Bu sınırlamayı aşmak ve dil işleme süreçlerinin hem işitsel hem de anlam temelli boyutlarını nörobilimsel olarak ayırt edebilmek amacıyla, bu araştırma yüksek zaman çözünürlüklü manyetik ensefalografi (MEG) verilerini kullanarak işitsel algı ile dilsel anlamanın nöral temellerini karşılaştırmayı amaçlamaktadır.

**Araştırma Soruları:** Çalışmada sorular açıkça belirtilmemiş olsa da temelde şu sorulara yanıt aradığı söylenebilir:

* MEG verileriyle konuşma temelli işitsel uyaranlara karşı beyin aktivitesi, ses temsilleri (örneğin zaman-frekans çözümlenmesi ve wav2vec2) kullanılarak ne ölçüde tahmin edilebilir?
* Aynı MEG verileri, metin temsilleri (örneğin CLIP ve GPT-2 gömme vektörleri) kullanılarak ses temsillerine kıyasla daha iyi tahmin edilebilir mi?
* İşitsel algı (ses temelli) ve dilsel anlama (metin temelli) süreçleri, beyinde hangi bölgelerde ve hangi frekans bantlarında farklı şekilde temsil edilmektedir?
* Beynin konuşma işleme sürecinde, akustik bilgi mi yoksa anlam odaklı dil bilgisi mi daha belirleyici rol oynamaktadır?

**Yöntem:** Bu araştırma, MEG-MASC adlı halka açık veri setinden elde edilen manyetik ensefalografi (MEG) verileri üzerinde yürütülmüştür. Çalışmaya, doğal hikâyeler dinlerken kaydedilen verileri içeren 8 katılımcı dahil edilmiştir. Ham MEG sinyalleri, 0.5–30 Hz bandında bant geçiren filtreleme, kelime başlangıçlarına hizalanmış 3 saniyelik pencerelere bölünme, 200 ms’lik temel çizgi düzeltmesi ve genlik kırpma gibi adımlarla önişlemeye tabi tutulmuştur. Araştırmada iki tür kodlama modeli (encoding model) geliştirilmiştir: (1) Ses-tabanlı modeller, giriş olarak zaman-frekans çözümlenmesi (STFT) ve wav2vec2’den elde edilen gizil temsilleri kullanırken; (2) Metin-tabanlı modeller, her MEG penceresine karşılık gelen 20 kelime öncesi bağlam ve 5 kelime sonrası olmak üzere toplam 25 kelimelik dilsel bağlamı CLIP ve GPT-2 dil modelleriyle gömme vektörlerine dönüştürmüştür. Tüm modellerde, bu temsiller ile MEG sinyalleri arasındaki ilişkiyi tahmin etmek için ridge regresyon uygulanmış, modeller %70 eğitim – %30 test şeklinde bölünen veriyle eğitilmiş ve çapraz doğrulama ile en iyi düzenlileştirme parametresi (λ = 5000) seçilmiştir. Model performansı, Pearson korelasyonu (PC) ve R² skorları ile değerlendirilmiş; istatistiksel anlamlılık, 100 kez rastgele karıştırma testi ile doğrulanmıştır. Analizler, tam frekans bandı (0.5–30 Hz) ve delta, teta, alfa, beta alt bantlarında ayrı ayrı yürütülmüştür.

**Bulgular ve Sonuçlar:** Araştırmanın temel bulguları, metin temsillerinin (CLIP ve GPT-2) MEG sinyallerini ses temsillerine (zaman-frekans çözümlenmesi ve wav2vec2) göre daha yüksek doğrulukla tahmin edebildiğini göstermiştir; bu, beynin konuşulan dili sadece akustik bir sinyal olarak değil, anlam taşıyan bir dil yapısı olarak işlediğine işaret etmektedir. Beyinsel dağılım açısından, ses temsilleri özellikle lateral temporal korteks gibi işitsel işlemlemeyle ilişkili bölgelerde, metin temsilleri ise frontal korteks (özellikle Broca bölgesi) gibi üst düzey dil ve anlam entegrasyonuyla ilişkili alanlarda daha güçlü korelasyonlar üretmiştir. Frekans analizleri, en yüksek kodlama performansının 8–30 Hz aralığında (alfa ve beta bantları) gerçekleştiğini ortaya koymuş; bu bantlar uyanıklık, dikkat ve bilişsel işlemeyle ilişkilidir. Tüm modellerin performansı istatistiksel olarak anlamlı bulunmuş (p < 0.05; Z-skorları 5.8–8.1 aralığında), rastgele etkilerin dışlandığı doğrulanmıştır.

**Tartışma ve Çıkarımlar:** Bu çalışma, işitsel dil uyaranlarının işlenmesinde işitsel algı ile dilsel anlamanın ayrı nöral yollarla temsil edildiğini göstermektedir. Bulgular, beynin konuşma sinyallerini sadece fiziksel ses özellikleriyle değil bağlamsal ve semantik içerikle birlikte işlediğini desteklemektedir; bu da metin temsillerinin (GPT-2 ve CLIP) MEG sinyallerini ses temsillerinden daha iyi tahmin etme başarısına yansımıştır. Beyinsel dağılım analizleri, şakak bölgelerinin (superior temporal gyrus) işitsel girdiyi, alın bölgesinin (frontal korteks) ise anlam entegrasyonu ve bilişsel kontrol süreçlerini üstlendiğini ortaya koymuştur. Özellikle 8–30 Hz frekans aralığında (alfa ve beta bantları) gözlenen güçlü korelasyonlar, dil işleme sırasında beynin aktif, dikkatli ve bilişsel olarak katılımlı bir durumda olduğunu göstermektedir. Bu bulgular, dilin nöral işlenmesinin hiyerarşik ve çok bileşenli bir süreç olduğunu vurgulamakta; duyusal girdi, anlam çıkarımı ve yürütücü işlevler bir arada devreye girmektedir.

Bu çalışmanın diğer bir sonucu da eğitilmiş derin öğrenme modellerinin (GPT-2, wav2vec2 vb.) bilişsel nörobilimde geçerli temsil araçları olarak kullanılabileceğini göstermektedir. Ancak bu modellerin kendi içsel önyargıları ve mimari sınırlılıkları dikkate alınmalıdır. Gelecek çalışmalarda, daha geniş ve çeşitli katılımcı grupları, çok modlu nörogörüntüleme verileri (MEG + fMRI gibi) ve daha gelişmiş dil modelleri ile bu bulguların genellenebilirliği sınanabilir. Bununla birlikte, bireyin zihinsel süreçlerini giderek daha hassas şekilde çözebilen bu teknolojilerin etik çerçeveler içinde geliştirilmesi zorunludur; özellikle beyin verilerinin mahremiyeti, bilişsel gizlilik ve kötüye kullanım riskleri dikkatle yönetilmelidir.

Sonuç olarak, bu çalışma, dilin nöral temellerini anlamada hesaplamalı modelleme ile nörogörüntülemenin güçlü bir sentezini sunmakta ve dil işleme mekanizmalarına dair hem teorik hem uygulamalı yeni yollar açmaktadır.

**Değerlendirme:** Yapılan değerlendirme sonucunda, incelenen çalışmanın hem güçlü hem de zayıf yönler taşıdığı belirlenmiştir.

Bu makale, dil işleme süreçlerini hesaplamalı nörobilim perspektifinden incelenmesi açısından literatüre önemli bir metodolojik ve kavramsal katkı sunmaktadır. Çalışmanın temel gücü, Manyetoensefalografi (MEG) verilerini kullanarak hem işitsel hem de anlamsal derin öğrenme temsillerini karşılaştırmasıdır. Özellikle MEG'in sunduğu yüksek zaman çözünürlüğü, fMRI'nın yetersiz kaldığı dil işlemenin dinamik nöral analizine olanak tanımakta ve yöntemsel bir ileri görüşlülük sergilemektedir. Ayrıca, araştırmanın klasik zaman-frekans analizi ile birlikte wav2vec2, CLIP ve GPT-2 gibi derin öğrenme modellerini aynı çerçevede sınaması, sadece 'hangi model en iyi tahmini yapar?' sorusunun ötesine geçerek, hangi bilgi türünün beynin hangi nölümünde kodladığına dair bilişsel bir sorgulama yapmaktadır. Bu kapsamlı model karşılaştırması ve delta, teta, alfa gibi frekans bantlarına dayalı detaylı analiz, işitsel ve semantik işleme arasındaki ayrışan nöral temelleri net bir şekilde ortaya koyarak çalışmanın bilimsel değerini yükseltmektedir. Son olarak, rastgele karıştırma (permutation) testleri, Z-skorları ve p-değerleri ile yapılan titiz istatistiksel doğrulama, elde edilen bulguların güvenilirliğini pekiştirerek akademik sağlamlığı desteklemektedir.

Bununla birlikte, çalışmanın yorumlanmasını ve genellenebilirliğini sınırlayan kritik metodolojik eksiklikler bulunmaktadır. Başta, sadece sekiz (8) katılımcı ile yürütülmüş olması, temel metodolojik zayıflığı oluşturmaktadır. Ayrıca bu katılımcı grubun demografik ve bilişsel profillerine dair çok fazla bilgi sunulmamıştır. İkinci olarak, çalışmanın sadece dinleme görevine odaklanan tek yönlü uyarım tasarımı, davranışsal verilerin (anlama kontrolü, üretme testi vb.) toplanmamış olması nedeniyle nöral tepkilerin gerçekten dilsel "anlam" işlenmesiyle mi ilişkili olduğunu davranışsal olarak doğrulamayı imkânsız kılmaktadır. Model seçimi konusunda da önemli bir dengesizlik mevcuttur. Şöyle ki, wav2vec2 bir konuşma modeli iken GPT-2 ve CLIP metin modelleridir. Bu durum, ses ve metin karşılaştırmasının yanı sıra modalite farkını da içermekte, dolayısıyla beyinde gerçekleşen kodlamadaki farklılığın anlamla mı ilgili yoksa sadece girdi formatındaki farktan mı kaynaklandığı sorusunu belirsizleştirmektedir.

Sonuç olarak çalışmanın derin öğrenme modellerinin bilişsel nörobilim araştırmalarında geçerli ve verimli araçlar olarak kullanılabileceğini göstermesi açısından önemlidir ve gelecek araştırmalar için sağlam bir metodolojik temel sunmaktadır. Ancak, katılımcı sayısı düşüklüğü ve farklı tür girdi sağlayan modellerin seçimi gibi mevcut sınırlılıklar, bulguların yorumlanmasına sınırlılık getirmektedir.