



seit 1558

Hörst du mein Gesicht? Neurophysiologische Korrelate multimodalen Stimmenlernens

Mareike Dietz & Christoph Roß

Leitung: Dr. Jürgen M. Kaufmann & Dr. Romi Zäske
Institut für Psychologie, Friedrich-Schiller-Universität Jena



Einleitung

Bekannte Personen können anhand ihres Gesichts und ihrer Stimme identifiziert werden. Der Zeitpunkt der Integration von Gesicht und Stimme wird momentan kontrovers diskutiert (s. Abb.1). So zeigten von Kriegstein et al. (2005), dass bei unimodaler Präsentation bekannter Stimmen auch gesichtsverarbeitende Areale aktiviert werden. Aufgrund der ungenauen zeitlichen Auflösung der verwendeten fMRT-Methode blieb jedoch offen, ob diese Aktivierung tatsächlich durch frühe multimodale Integration oder aber Top-Down-Prozesse zustande kam. In der vorliegenden Studie nutzten wir die hohe zeitliche Auflösung ereignis-korrelierter Potentiale (EKPs) zur Untersuchung obiger Fragestellung. Basierend auf Befunden zur verbesserten Lernleistung für markante gegenüber gewöhnlichen Gesichtern (Kaufmann & Schweinberger, in Revision), untersuchten wir, ob auch Stimmen besser wiedererkannt werden, wenn sie zuvor zusammen mit markanten statt mit gewöhnlichen Gesichtern gelernt werden. Frühe Unterschiede (< 300 ms) für in der Testphase unimodal dargebotene Stimmen, die zuvor in der Lernphase entweder mit markanten oder mit gewöhnlichen Gesichtern gepaart präsentiert wurden, sprächen für eine frühe Integration multimodaler Personeninformation.

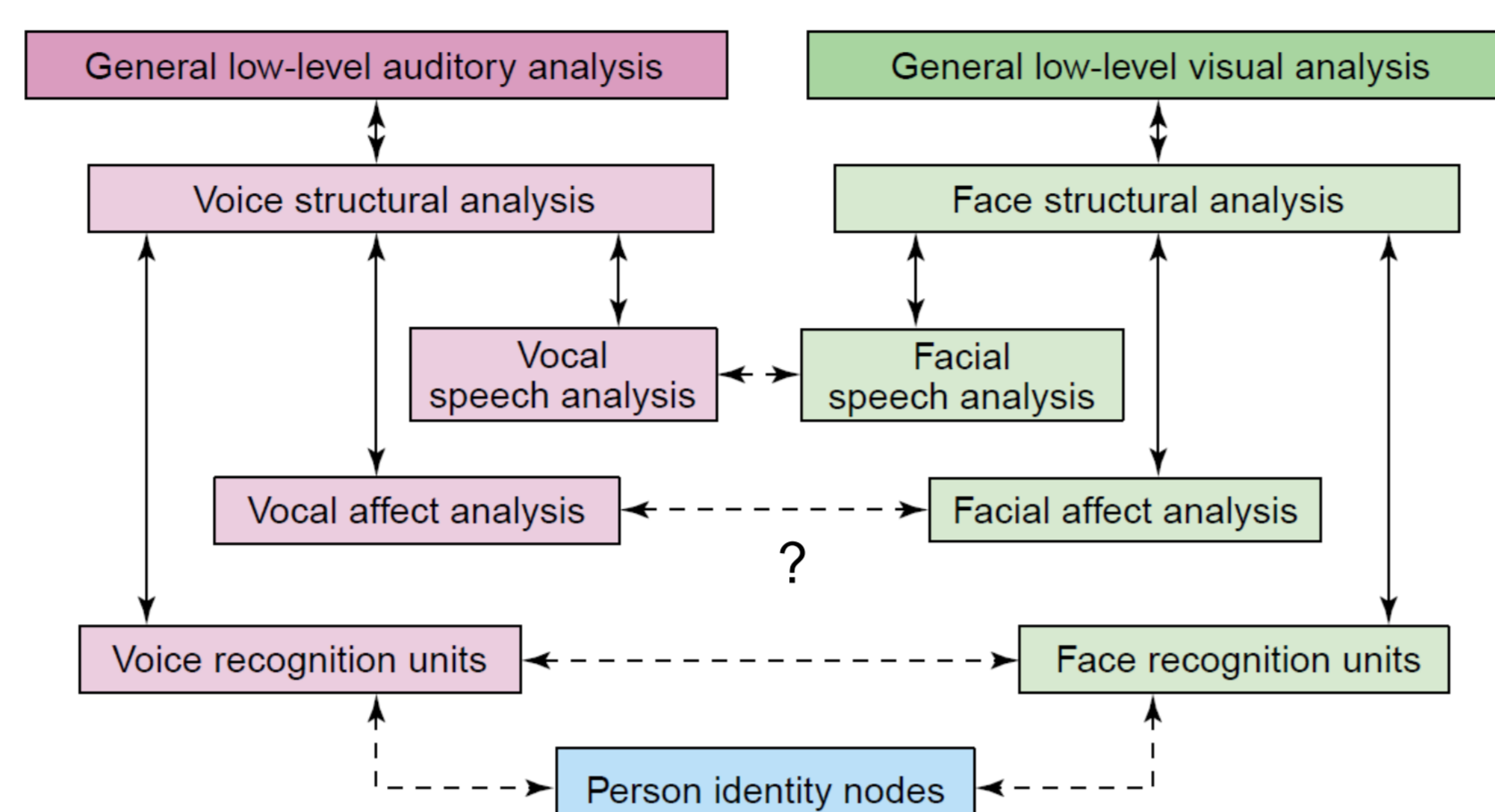


Abb. 1: Modell der Stimmen- und Gesichterwahrnehmung mit möglicher früher Interaktion („?“) nach Belin et al. (2004)

Methoden

Stichprobe

- 20 Versuchspersonen (VPs) im Alter von 19 bis 28 Jahren (18 Frauen, mittleres Alter 22 Jahre)

Stimuli

- Stimmproben von 24 Sprechern (12 Frauen) im Alter von 18-22 aus einer Datenbank der DFG Forschergruppe „Person Perception“
- 3 unterschiedliche Sätze pro Sprecher (Sampling Rate 44.1 kHz, Auflösung 16 Bit) mit spitzennormalisierter Lautstärke
- 12 Gesichter (6 Frauen) aus der Datenbank des Lehrstuhls auf grauem Hintergrund, von denen in einer Pilotstudie jeweils die Hälfte als besonders markant oder als eher gewöhnlich eingeschätzt wurde

Prozedur (s. Abb. 2)

- Probanden durchliefen 6 Blöcke mit jeweils einer Lern- und Testphase
- Lernphase:** 12 Gesichter (6 markant, 6 gewöhnlich) wurden randomisiert nacheinander präsentiert, zusammen mit jeweils zwei verschiedenen Sätzen eines Sprechers gleichen Geschlechts; die Probanden wurden vorher aufgefordert, sich die Sprecher einzuprägen
- Testphase:** Probanden hörten einen 3. Satz, gesprochen von 24 unterschiedlichen Sprechern (12 aus Lernphase, 12 unbekannt) und entschieden, ob die Stimme aus der Lernphase bekannt war oder nicht
- 2 Ausbalancierungsbedingungen bzgl. der Stimmen:
 - (1) Verwendung der Stimme in Lernphase oder nicht
 - (2) Paarung der zu lernenden Stimme mit markantem oder gewöhnlichem Gesicht> Zuordnung Gesichter zu Stimmen blieb innerhalb einer VP konstant

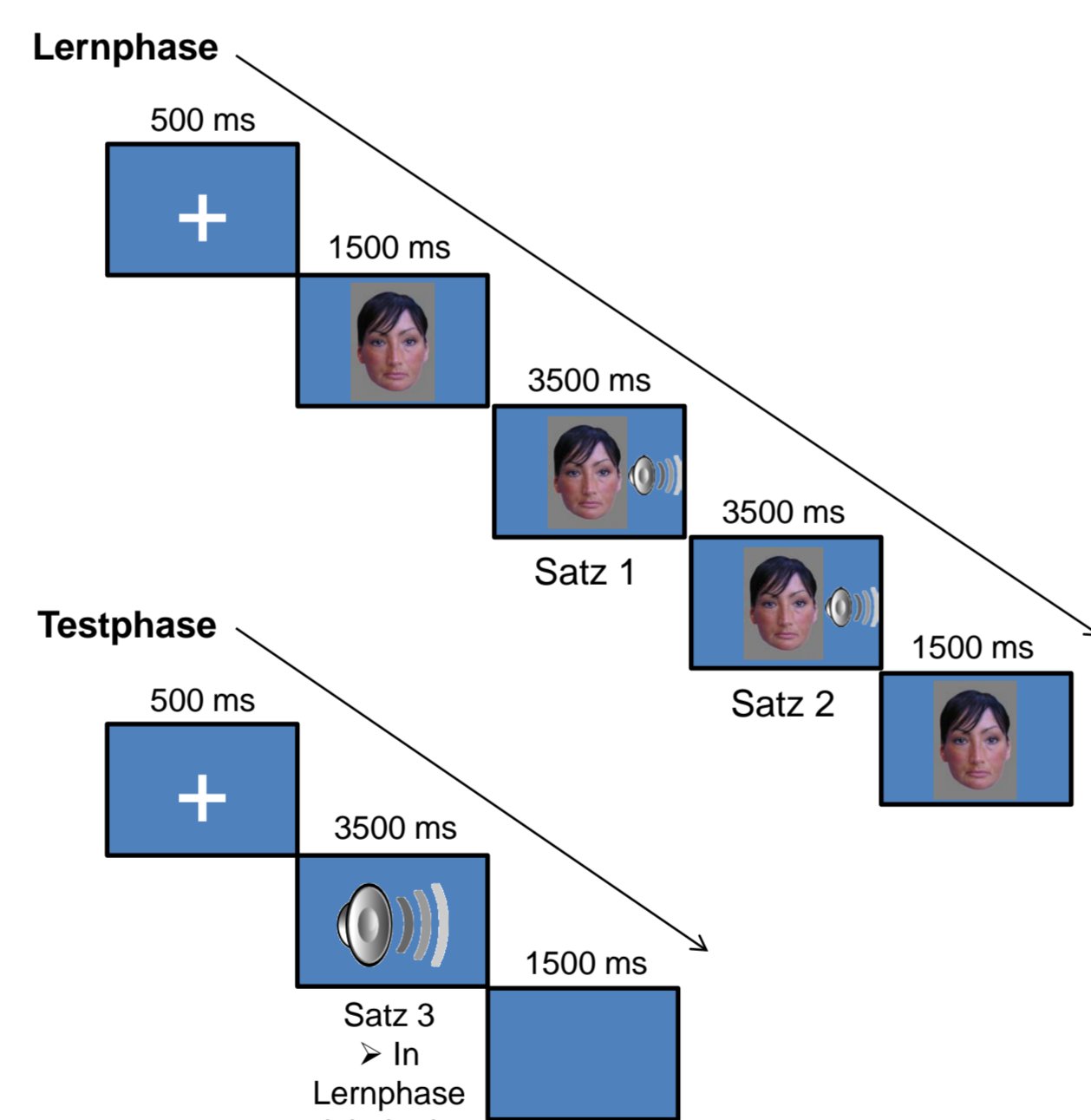


Abb. 2: Schematischer Ablauf der Lern- und Testphase mit Beispiel für ein markantes Gesicht

EEG – Aufzeichnung und Analyse

- Ableitung von 32 Elektroden (AC, 0.05 Hz Hochpass, 100 Hz Tiefpass, Abtastrate 500 Hz, Referenz Cz)
- okulare Artefakte wurden nach Erhebung mit BESA (Version 1.8.10) korrigiert
- nach Mittelung wurden ereignis-korrelierte Potentiale (EKPs) gefiltert (20 Hz Tiefpass) und auf Durchschnittsreferenz umgerechnet

- Berechnung durchschnittlicher Amplituden für die EKP-Komponenten der Stimmen in Lern- und Testphase (N1: 120-160 ms, P2: 210-250 ms, jeweils an Cz; „N250“ an PO9/PO10; LPC: 600-1000 ms an Pz; 750-950 ms an O1/O2, P7/P8, P9/P10, PO9/PO10) sowie der EKPs für Gesichter während der Lernphase (N170: 146-186 ms; P200: 230-280 ms; N250: 280-380 ms, jeweils an P9/P10 und PO9/PO10; LPC: 400-600 ms an Cz)

Literatur

- Belin, P., Fecteau, S. & Bédard, C. (2004). Thinking of the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, 8,129-135.
- Kaufmann, J. M. & Schweinberger, S. R. The faces you remember: Caricaturing shape in photographic images facilitates brain processes reflecting the acquisition of new face representations. (in Revision für *Biological Psychology*).
- Puce, A., Allison, T., Asgari, M., Gore, J.C., McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience*, 16, 5205-5215.
- von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, 17, 367-376.
- Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing facial identities. *Quarterly Journal of Experimental Psychology*, 60, 1446-1456.

Ergebnisse

Verhaltensdaten (s. Abb. 3):

ANOVAs mit Messwiederholung auf dem Faktor Lernbedingung (dreistufig: „mit markantem“ vs. „mit gewöhnlichem Gesicht gelernt“ und „nicht gelernt“) für Reaktionszeiten (RTs) und Antwortgenauigkeiten bei der Stimmenklassifikation; zusätzlich Testung orthogonaler Kontraste für die Bedingungen *markant* und *gewöhnlich*

> sign. Haupteffekt der Lernbedingung für RTs, $F[1,6,30.2] = 6.04, p < .05$ (df korrigiert nach Huynh-Feldt); sign. Mittelwertsunterschied zwischen *markant* und *gewöhnlich*, $t[19] = -2.38, p < .05$

> sign. Haupteffekt der Lernbedingung für Antwortgenauigkeiten, $F[2,38] = 8.21, p < .01$; kein sign. Mittelwertsunterschied zwischen *markant* und *gewöhnlich*, $t[19] = 0.93, p = .36$

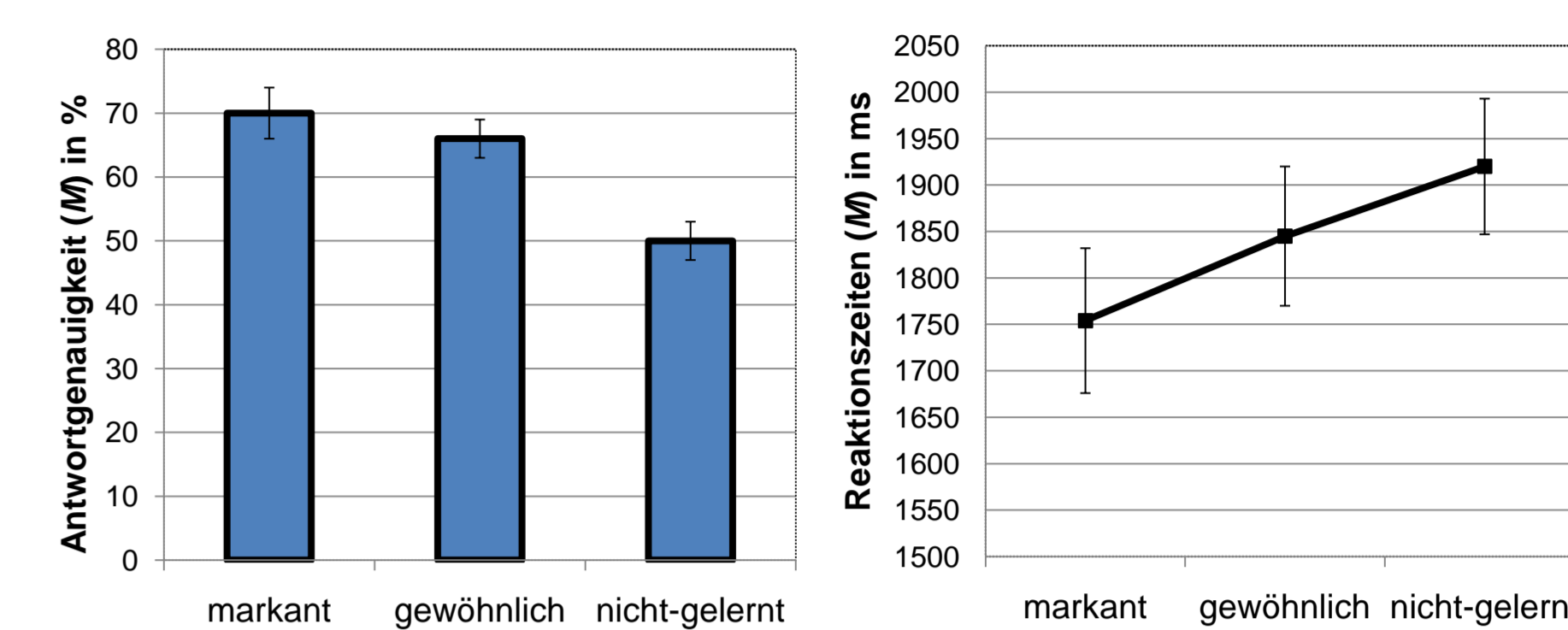


Abb. 3: mittlere Reaktionszeiten und Antwortgenauigkeiten pro Lernbedingung (Testphase) mit Standardfehlern der Mittelwerte

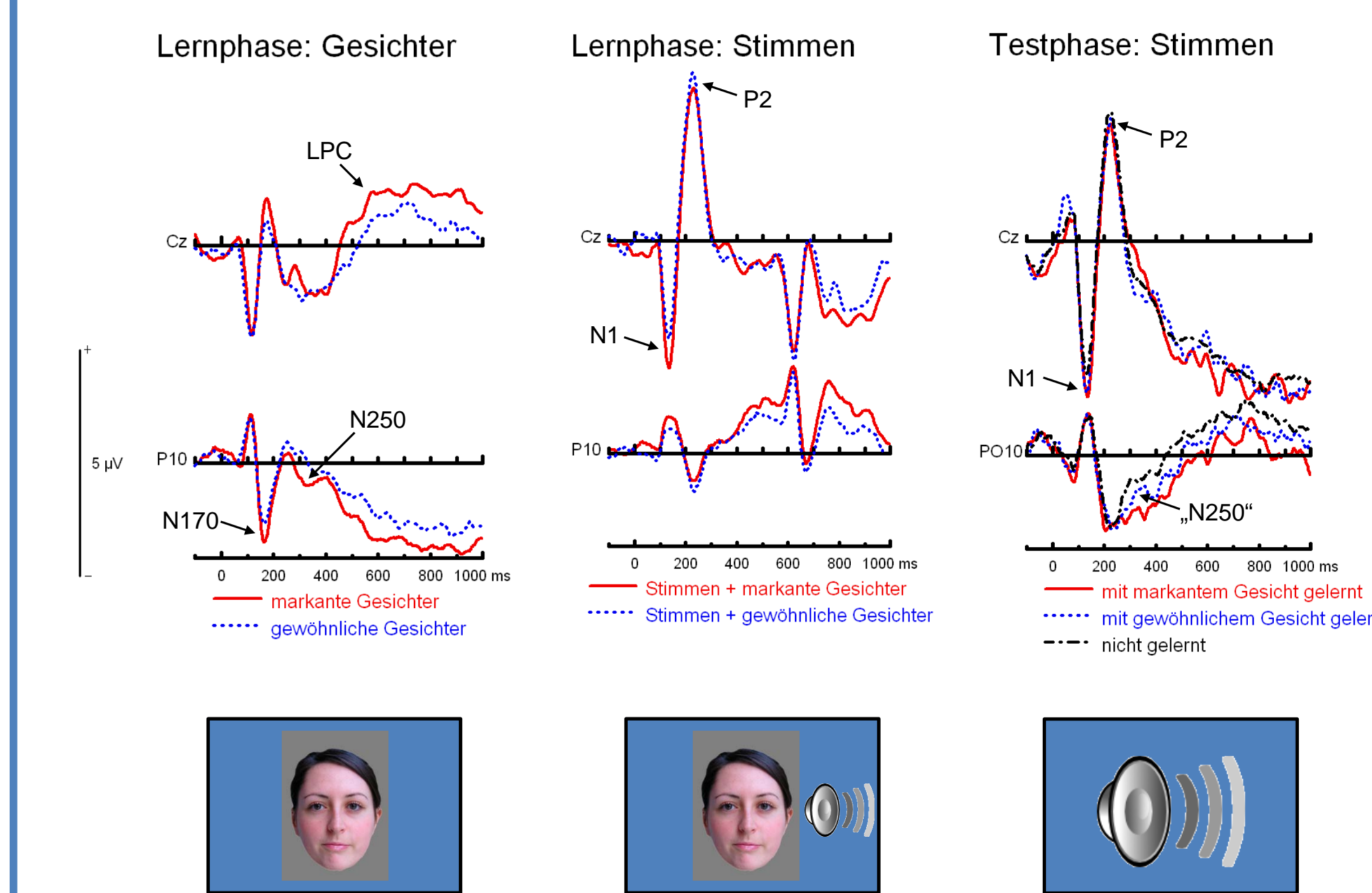


Abb. 4: Grand Mean Averages (GMAs) der ereignis-korrelierten Potentiale pro Bedingung und pro Experimentalphase mit Beispiel für ein gewöhnliches Gesicht

EKPs (s. Abb. 4):

Lernphase:

- N250: größere Amplitude für gewöhnliche vs. markante Gesichter, $F[1,19] = 5.03, p < .05$
- N1 (120-160 ms): größere Amplitude für Stimmen mit markanten Gesichtern, $t[19] = -2.76, p < .05$
- 750-950 ms: stärkere Positivierung okzipito-temporal für mit markanten Gesichtern gelernte Stimmen, $F[1,19] = 18.17, p < .001$

Testphase:

- „N250“: sign. Haupteffekt des Faktors Lernbedingung, $F[2,38] = 3.27, p < .05$
 - sign. linearer Trend mit zunehmender Negativierung von nicht-gelernt über mit gewöhnlichem Gesicht zu mit markantem Gesicht gelernt, $F[1,19] = 7.85, p < .05$
 - kein sign. Unterschied zwischen *markant* und *gewöhnlich*, $F[1,19] = 1.14, p = .29$, aber zwischen *markant* und *nicht gelernt*, $F[1,19] = 7.85, p < .05$.

Quellenlokalisierung (s. Abb. 5)

Für die Testphase ergab eine Quellenlokalisierung (Besa™, 5.1.8.10) der auditorisch evozierten Potentiale zusätzliche Evidenz für eine mögliche frühe Aktivierung von Gesichterrepräsentationen durch Stimmen: für die Differenz zwischen mit markanten Gesichtern gelernten und nicht gelernten Stimmen wurde für den Zeitbereich der N250 eine Quelle im fusiformen Gyrus ermittelt. Hierbei handelt es sich um eine Hirnstruktur, die mit der Verarbeitung von Gesichtern in Verbindung gebracht wird (siehe z.B. Puce et al., 1996).

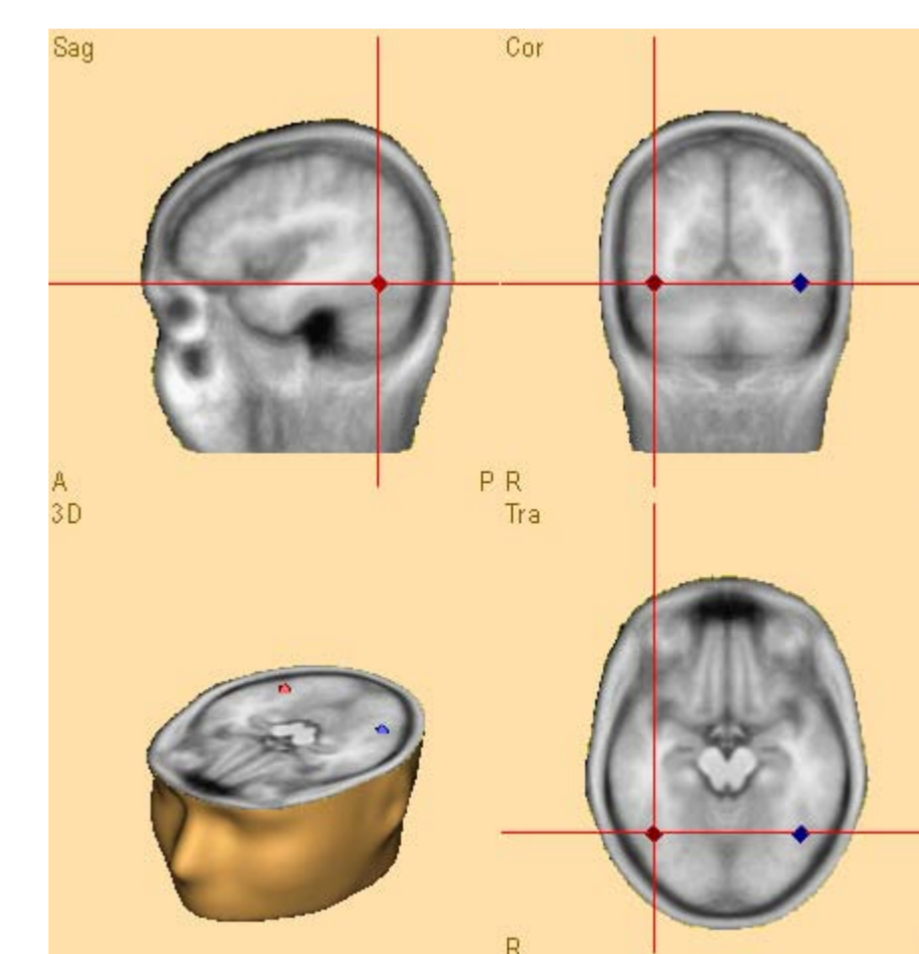


Abb. 5: Quellenlokalisierung der N250 auf Stimmen (Differenz gelernt mit markanter Stimme vs. nicht gelernt) projiziert auf Standard MRI (Talairach-Koordinaten: x = ±42, y = -59, z = -12).

EKP Unterschiede

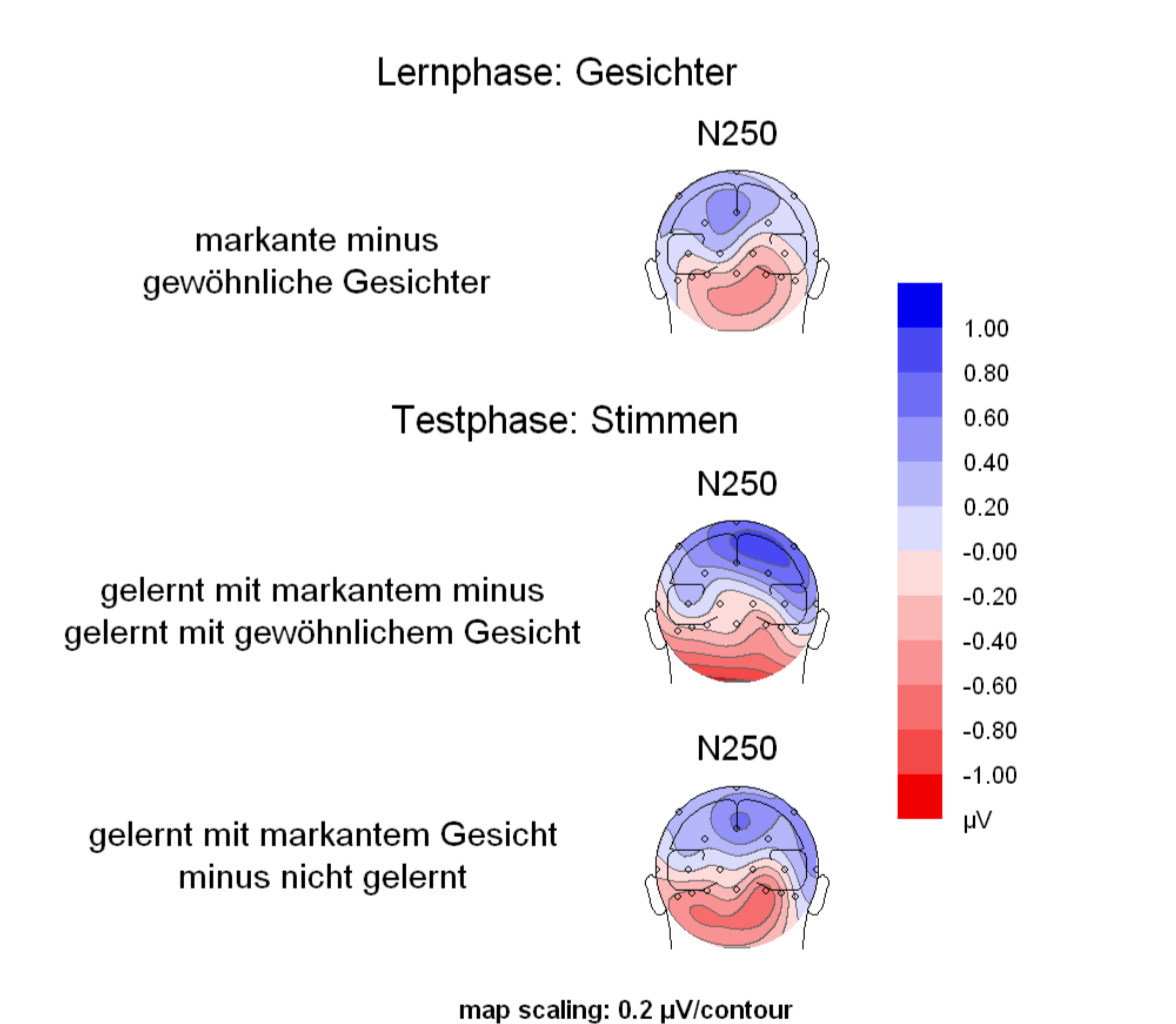


Abb. 6: Topographie der Unterschiede (spherical spline interpolation) für den Zeitbereich der N250 (280-380 ms)

Diskussion

- Evidenz für frühe Verarbeitungsunterschiede bei Stimmen in Abhängigkeit von Charakteristika des Gesichts (hier Markantheit)
 - effizienteres Lernen von Stimmen mit markanten Gesichtern
 - spricht für multimodale Aktivierung bei der Personenwahrnehmung
- EKPs: unimodale Präsentation einer Stimme kann die zugehörige Gesichtsrepräsentation bottom-up aktivieren
 - Unterschiede bei „N250“ unterstützen die Annahme einer frühen Interaktion der Gesichts- und Stimmenverarbeitung
 - zusätzliche Evidenz: Quellenlokalisierung erweitert Befunde von von Kriegstein et al. (2005) um die zeitliche Dimension
- Manipulation von Gesichtermarkantheit war erfolgreich: EKP Unterschiede auf Gesichter in Lernphase
- Grenzen:
 - Wiedererkennungsleistung für Stimmen trotz sechsmaliger Wiederholung insgesamt eher schlecht (aber über Zufallsniveau, $d' > 0$)
- Empfehlung für zukünftige Forschung:
 - Verwendung von Videos statt statischer Bilder für höhere ökologische Validität (siehe z.B. Schweinberger et al. 2006)