

Audio-visuelle Wahrnehmung von Vokalen im Deutschen, Ungarischen, Georgischen und Ägyptisch-Arabischen

Winter, Isa Samira¹; Juhász, Kornélia²; Deme, Andrea²; Greisbach, Reinhold¹

¹IfLPhonetik, Universität zu Köln, ²Dept. Phonetics, Eötvös Loránd University Budapest

Phonetische Untersuchungen fokussieren in der Regel nur auf das Sprachsignal. Findet die lautsprachliche Kommunikation jedoch wie üblich in einer Angesicht-zu-Angesicht-Situation zwischen Sprecher und Hörer statt, steht zusätzlich auch noch der optische Informationskanal für den Hörer zur Verfügung, um die vom Sprecher gesendete Information zu dekodieren. In dieser Studie wird die Auswirkung dieser zusätzlichen Information auf die Sprachwahrnehmung untersucht.

Speziell untersucht diese Studie wie sich die zusätzlich zum Sprachsignal sichtbare Form der Lippen auf die Wahrnehmung des linguistischen Merkmals Lippenrundung bei Vokalen auswirkt. Dabei wird dieser Effekt sprachvergleichend jeweils zwischen zwei Sprachen untersucht, deren Lautsystem über gerundete vordere Vokale verfügt (Deutsch und Ungarisch; Wahrnehmungskontrast /e/ vs. /ø/) und zwei Sprachen, bei denen dies nicht der Fall ist (Georgisch und Ägyptisch-Arabisch; Wahrnehmungskontrast /e/ vs. /o/).

Die Perzeptionsexperimente fanden online in den Heimatländern der Teilnehmer statt. Für jede Sprache nahmen jeweils 20 Muttersprachler der jeweiligen Sprachen an der Studie teil.

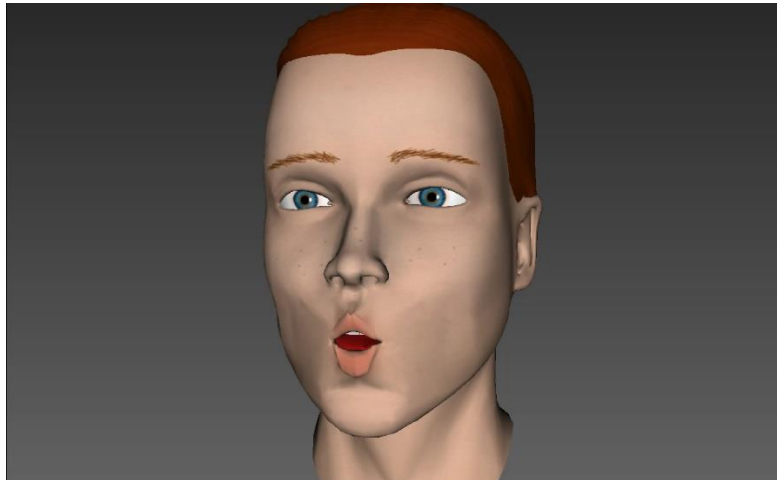
Für die akustischen Stimuli wurden ausgehend von einem durch einen deutschen männlichen Sprecher gesprochenen [e:] mit Hilfe der in PRAAT [1] implementierten Formantsynthese künstliche Vokale erzeugt. Dabei wurde jeweils der F2 (von 2100 Hz) in mehreren Zwischenschritten so weit erniedrigt, dass zunächst die Klangfarbe eines [ø:] (bei ca. 1700 Hz) und durch eine weitere Verringerung die Klangfarbe eines [o:] (Minimum bei 600 Hz) entstand. Für das Experiment im Deutschen und Ungarischen wurden nur die Signale mit F2 zwischen 1450 – 2100 Hz verwendet, für das Experiment im Georgischen und Ägyptisch-Arabischen der gesamte Bereich von 600 – 2100 Hz. Zur Erzeugung der zusätzlichen visuellen Stimuli wurde das Talking Head System MASSY von Fagel u.a. ([2],[3]) benutzt, mit dem jeweils ein dynamisches (von Ruheposition zum Vokal und wieder zurück) [e:] (ungerundete Lippen) und [o:] bzw. [ø:] (gerundete Lippen; Abb. 1) erzeugt wurde.

Die künstlichen akustischen Vokalsignale wurden einmal ohne zusätzliches optisches Signal und einmal mit einem visuellen [e:] sowie einem visuellen [o] bzw. [ø:] zusammengebunden und in zwei Telexperimenten, zunächst nur auditiv und anschließend audio-visuell randomisiert angeordnet. Die Teilnehmer hatten dann in einem Forced-Choice-Experiment die Aufgabe die Stimuli als [e:] oder [ø:] bzw. [e:] oder [o:] zu klassifizieren.

Als Resultat zeigt sich, dass bei Verringerung des F2 in allen Sprachen ein kontinuierlicher Übergang in der Wahrnehmung vom nicht-gerundeten Vokal [e:] zum gerundeten Vokal [ø:] bzw. [o:] stattfindet. Dabei verschiebt sich der 50%-Übergangspunkt der Wahrnehmung zwischen ungerundetem und gerundetem Vokal durch die zusätzliche optische Information in der erwarteten Art und Weise: Die sichtbaren gerundeten Lippen schieben den Übergangspunkt auf der F2-Skala nach oben, die sichtbaren ungerundeten Lippen ziehen ihn nach unten.

Allerdings gibt es Unterschiede in der Lage der Übergangspunkte, dem Grad der Verschiebung und sogar in der Form der Übergangskurve zwischen den jeweiligen Sprachpaaren (Deutsch und Ungarisch sowie Georgisch und Ägyptisch-Arabisch). Mögliche Ursachen für die sprachspezifischen Unterschiede in der Wahrnehmung werden diskutiert.

Abb. 1. MASSY_gerundetes Lippenbild



- [1] Boersma, P. & Weenink, D. Praat: doing phonetics by computer. *Computer program*, 2006.
- [2] Fagel, S. & Clemens, C.: An articulation model for audiovisual speech synthesis - Determination, adjustment, evaluation. *Speech Communication*, 44, 141-154, 2004.
- [3] Fagel, S. & Sendlmeier, W.F.: Das audiovisuelle Sprachsynthesystem MASSY – Implementierung und Optimierung. In: Beiträge zur 14. Konferenz Elektronische Sprachsignalverarbeitung, ESSV, Karlsruhe, 234-241, 2003.