Sprachproduktionstheorien (3)

Die kognitive Repräsentation phonetischer Ziele

Is speech planned in terms of

(1) auditory/acoustic goals? (Guenther, Perkell, et al.)

or

(2) vocal tract constrictions/gestures? (Fowler, Browman & Goldstein).

Evidence for (1) from Guenther et al. (1999)

"Articulatory tradeoffs reduced acoustic variability during American English /r/ production"

JASA 105(5), 2854-2865

American-English /r/

Artikulatorisch sehr variabel: Hauptvarianten ‘**bunched**’ (dorsal),
und **retroflex**

Aber akustisch ein sehr konstantes Merkmal: F3 sehr tief

Übungsaufgabe:

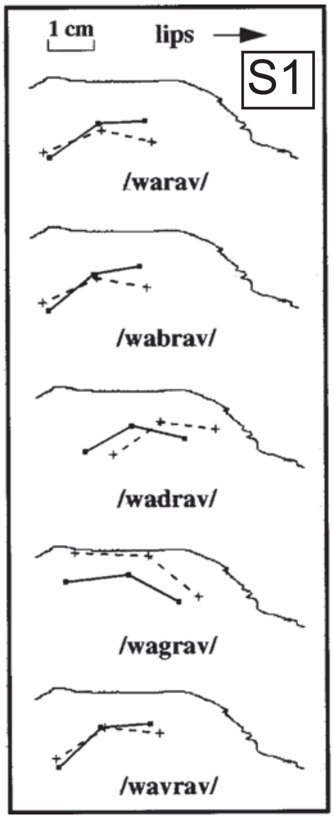
/r/ wurde in Logatomen in 5 verschiedenen Lautkontexten aufgenommen:

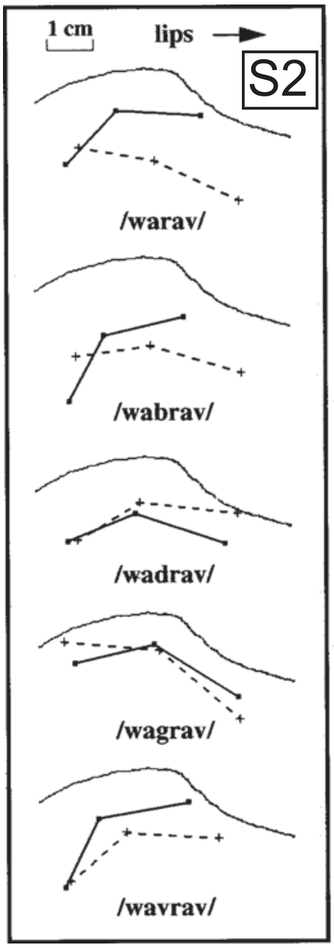
warav, wabrav, wadrav, wagrav, wavrav

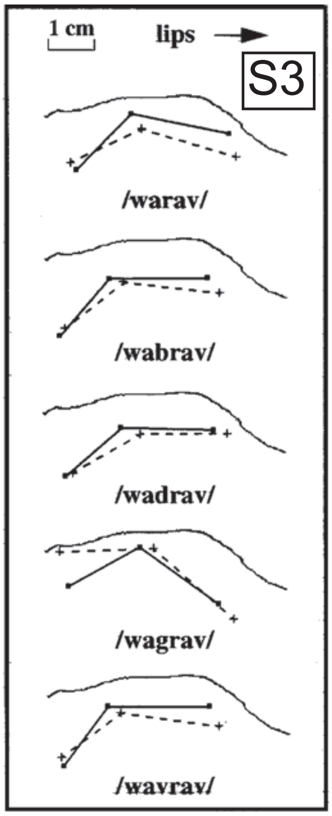
In welchen Kontexten ist die dorsale Variante besonders häufig zu finden?

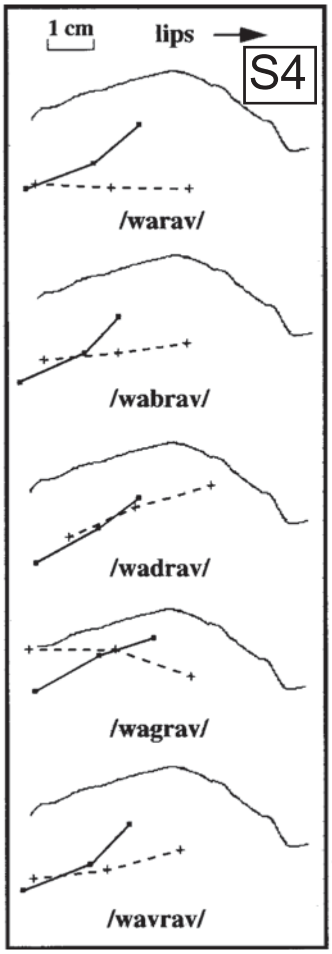
In welchen Kontexten die retroflexe Variante?

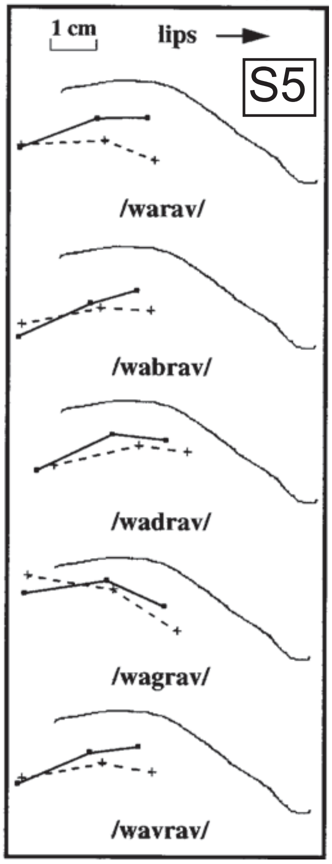
Gibt es Sprecher, die nur dorsal oder nur retroflex artikulieren?

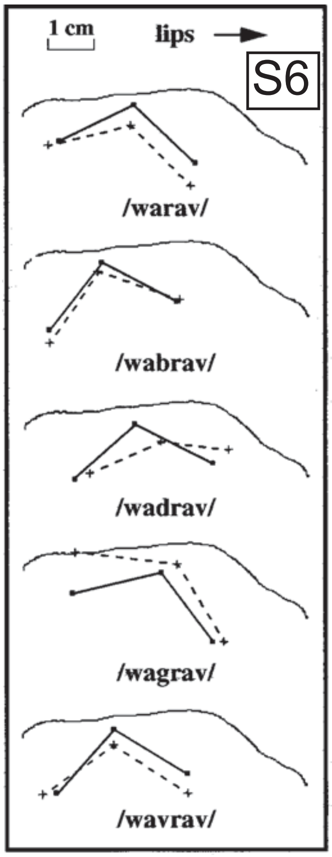


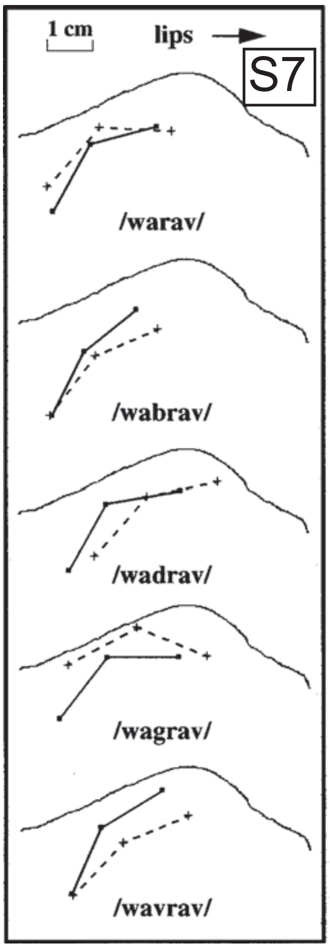












Guenther et al., 1999, Figs. 1 to 7.
Sample lingual articulations used by subjects 1 to 7 to produce /r/ in the five phonetic contexts. For each context, two schematized tongue shapes and a palatal trace are shown. Each tongue shape schematic was formed by connecting the three tongue transducers with straight lines. The tongue shape at the *F*3 minimum for /r/ is drawn with solid lines. The tongue shape 70 ms prior to the *F*3 minimum is drawn with dashed lines.

Folgende artikulatorische Einstellungen können alle zum tiefen F3 beitragen:

(1) je länger der vordere Hohlraum (front cavity)

(= je weiter hinten die Artikulationsstelle)

(2) je länger die Verengung an der Artikulationsstelle

(3) je enger die Verengung

desto tiefer F3

(festgestellt durch akustische Simulationen mit einem Vokaltraktmodell)

====> Durch Wechselspiel (tradeoffs) zwischen diesen 3 Einstellungen kann ähnlich tiefer F3 mit sehr unterschiedlichen Artikulationen erreicht werden.

Beispiel

Bei /r/ in Lautkontext A ist die Artikulationsstelle weiter vorne (vorderer Hohlraum kürzer) als in Lautkontext B.

Die resultierende Erhöhung von F3 kann ausgelichen werden, wenn in Kontext A die Verengung länger oder enger ist als in Kontext B.

Noch ein Blick auf die Rohdaten:

Wie unterscheiden sich die dorsalen und retroflexen Varianten in Hinblick auf diese 3 Einstellungen?

Kommen diese Wechselbeziehungen aber tatsächlich vor?

- EMA-Daten
- Viele Wiederholungen von /r/ in den 5 Lautkontexten.
- Analyse der Korrelationen zwischen den Mechanismen 1 bis 3.

Für alle 7 VPn:

- Deutliche tradeoff-Beziehung zwischen Länge des vorderen Hohlraums und Länge der Verengung

Für 6 von 7 VPn eine weitere deutliche tradeoff-Beziehung:

- 5x Länge der Verengung vs. Enge der Verengung
- 1x Länge des vorderen Hohlraums vs. Enge der Verengung

====>F3 wäre viel variabler, wenn Sprecher diese möglichen Wechselbeziehungen nicht ausnutzen würden

bestätigt durch Regressionsanalysen: Schätzung von F3 anhand der Artikulationsdaten

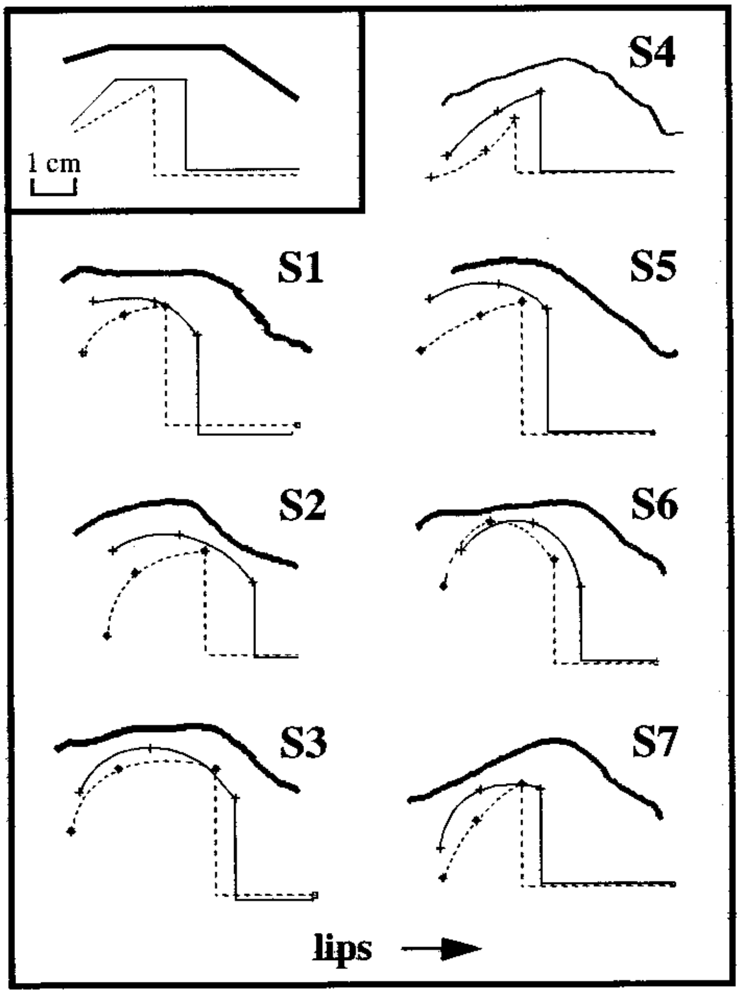


FIG. 17. Trading relations during /r/ production. The upper left corner shows two superimposed, highly schematized vocal tract outlines (dashed and solid lines) illustrating trading relations between front cavity length and palatal constriction length and area. Also shown are vocal tract outlines that illustrate the range of tongue shapes used by each of the seven subjects to produce /r/ in different phonetic contexts. Thin solid lines correspond to the tongue shapes for /r/ in /wagrav/ (averaged across repetitions), and dashed lines correspond to the /r/ in /wabrav/ or /warav/, depending on the subject. Thick solid lines indicate palatal outlines. Each outline is formed by connecting the three tongue transducer positions with a smooth curve, then projecting downward and forward from the frontmost tongue tongue transducer to the lower incisor transducer. All seven subjects show tradeoffs between the front cavity length and the constriction length and/or area when producing /r/ in the two different contexts.

Zitat Guenther et al. (1999), pp. 2863-2864:

“The issue of articulatory trading relations in speech production is relevant to current theories concerning the control of speech movements. Roughly speaking, computational models of speech motor control can be classified according to the type of phonemic ‘‘targets’’ that they use to command movements of the speech articulators. One type of computational model, exemplified by the task-dynamic model of Saltzman and Munhall (1989), utilizes a target for each phoneme that specifies important aspects of the shape of the vocal tract for that phoneme. This ‘‘vocal tract shape target’’ view is closely related to theories of speech perception and production in which the articulatory gesture serves as the basic unit of speech. These include the motor theory of speech perception (Liberman and Mattingly, 1985; Liberman et al., 1967), the direct realist theory of speech perception (Fowler, 1986, 1996), and the linguistic-gestural theory of phonology (Browman and Goldstein, 1990a,b).

A second type of computational model, exemplified by the DIVA model of speech acquisition and production (Guenther, 1995; Guenther et al., 1998), utilizes only an acoustic or auditory target for each phoneme, with no explicit vocal tract shape target. These models may use different shapes of the vocal tract to produce the same acoustic signal for a phoneme depending on things like phonetic context. Theories related to this ‘‘auditory target’’ view have been posited by various researchers (for some recent examples, see Johnson et al., 1993; Perkell et al., 1993, 1995, 1997; Savariaux et al., 1995a,b) and many of the roots for this line of thinking can be traced to the influential work of Jakobson et al. (1951).

A major difference between the auditory target and vocal tract shape target computational model classes is that the former explicitly predict the existence of articulatory trading relations when producing the same phoneme in different contexts, whereas the latter do not. Because the current results show the existence of trading relations in all seven subjects, they appear to favor acoustic target models over vocal tract shape target models. A potential reason for the use of articulatory tradeoffs is that they can reduce the amount of effort required to move the articulators through a set of acoustic targets. For example, the tongue shapes for /r/ in /wagrav/ were generally closer to the tongue shapes for /g/ than the tongue shapes for /r/ in other contexts, suggesting that, to a first approximation, subjects moved to the closest vocal tract shape that could be used to produce the appropriate sound in the prevailing phonetic context. The acoustic target control scheme used by the DIVA model has this property."

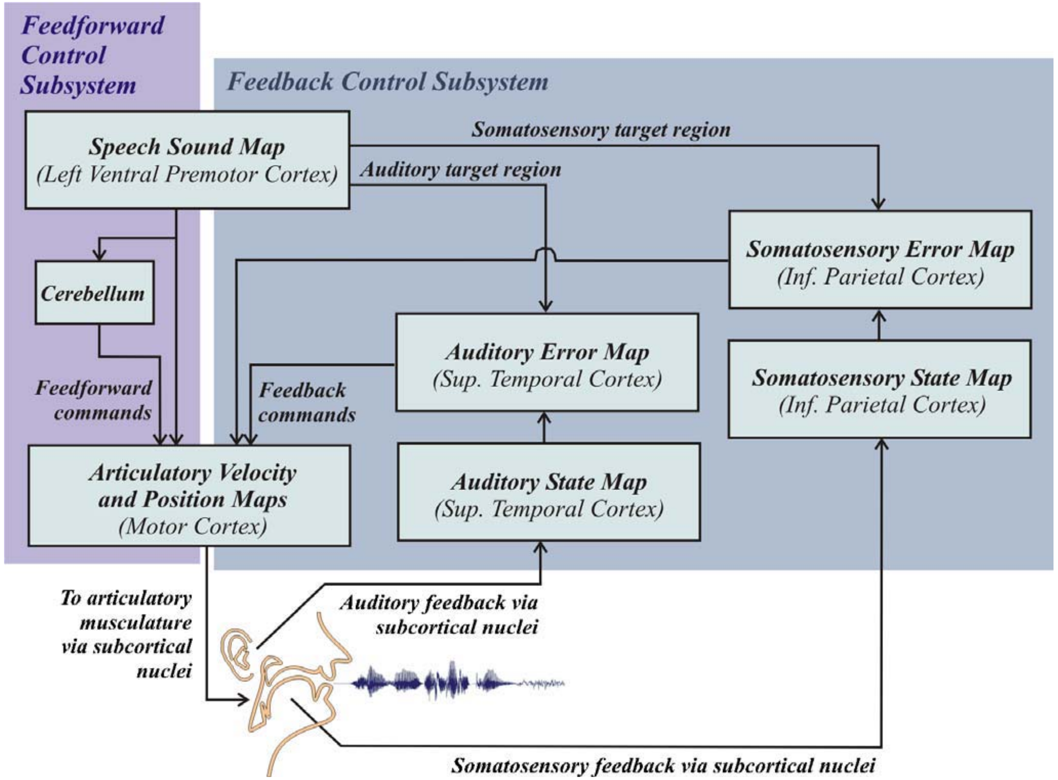


Fig. 1. Hypothesized neural processing stages involved in speech acquisition and production according to the DIVA model. Projections to and from the cerebellum are simplified for clarity.

DIVA demos:

http://www.bu.edu/speechlab/research/the-diva-model/model-demos