

Training of English prosody with acoustically modified voices

Radek Skarnitzl and Tomáš Bořil

Charles University (Prag)

Ricardo Urban

Aufbau

- Ziel der Studie
- Theoretischer Hintergrund Prosodie
- Ablauf der Studie
- Ergebnisdarstellung
 - Wahrgenommene Kompetenz
 - Analyse des Phrasings
 - Analyse der Produktion
- Diskussion

Ziel der Studie

- Es wurde ein Prosodietraining für Lernende des Englischen entworfen
- Es wurde mit verschiedenen Methoden überprüft, ob das Training erfolgreich ist

Prosodie

- Prosodie meint die **lautlichen Merkmale** der gesprochenen Sprache, die über einzelne Laute hinausgehen
- Dazu gehören z.B. Intonation, Betonung, Rhythmus und Phrasing
- Diese Merkmale können u.a. die grammatikalische Struktur des Satzes signalisieren, wichtige Wörter hervorheben oder Emotionen des Sprechenden anzeigen
- Manche der Funktionen sind universal, andere sind sprachspezifisch
- Richtiges Verwenden von prosodischen Merkmalen führt dazu, dass der Sprechende ähnlicher zu MuttersprachlerInnen klingt

Prosodie

- PSOLA-Verfahren (Pitch-Synchronous Overlap-and-Add) -> Tonhöhe und Sprechgeschwindigkeit modifizieren
- Pitch-Marks werden bestimmt
- Um den Pitch-Mark herum wird ein Stück des Signals ausgeschnitten
- Diese Segmente werden neu positioniert oder überlappend addiert

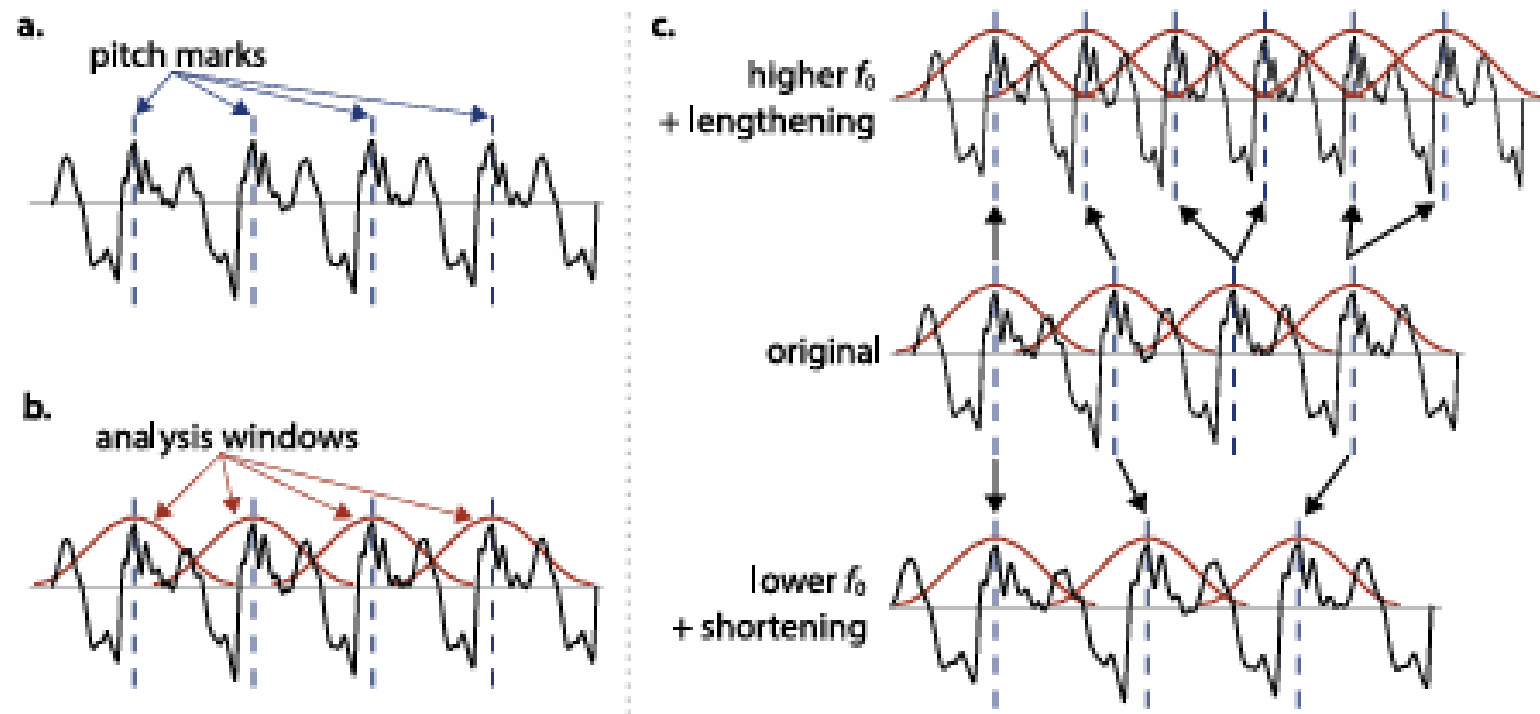


Figure 1. Block scheme of PSOLA manipulations: Identification of pitch marks (a); overlapping fundamental periods with windowing functions (b); and changing overlap to modify f_0 and adding or removing fundamental periods to modify duration (c)

Training

- 12 Teilnehmende (8 Frauen, 4 Männer)
- Tschechische MuttersprachlerInnen, die Englisch auf B1 bis B2 Niveau beherrschen
- Ein 230 Wörter langer Text wird von den Teilnehmenden als Vorbereitung eingesprochen
- Teile des Textes (6-10 Wörter lang) werden nach dem **PSOLA-Verfahren** manipuliert

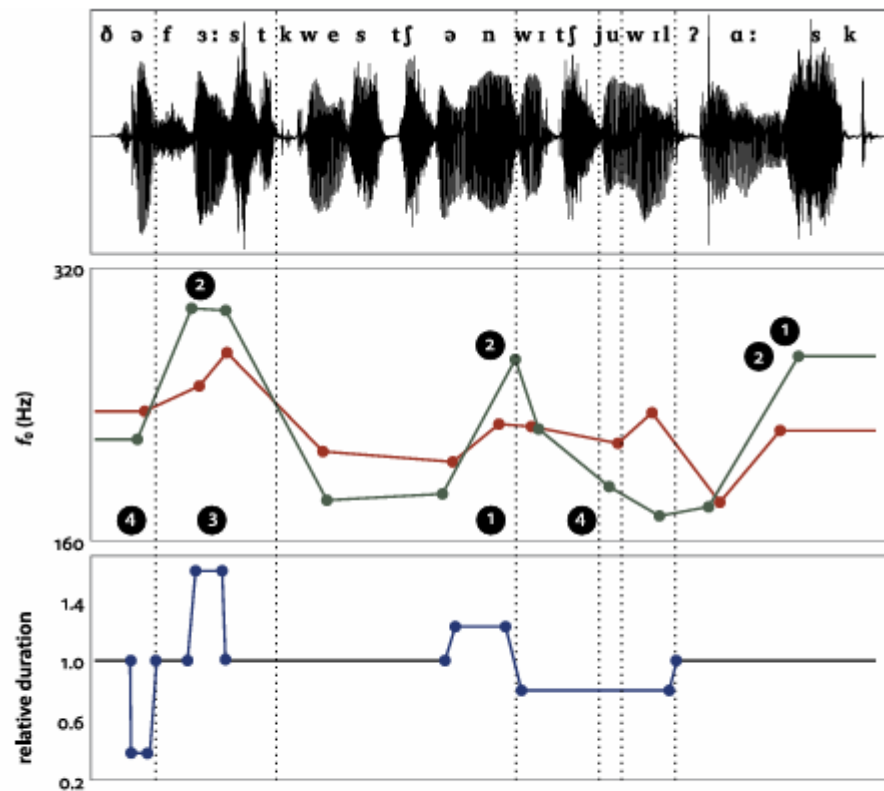


Figure 3. Examples of PSOLA manipulations on the portion *the first question which you will ask*: More salient phrasing signals (1), increased pitch range (2), prominence marking (3), shortening of unstressed syllables (4). The original melodic contour in red, the manipulated one in green

Training

- Das Training wird auf zwei Termine aufgeteilt
- 1. Trainingstermin:
 - 20 min Präsentation über Prosodie mit Übungen
 - Teilnehmende hören die modifizierte Audiodatei
 - 20 sek Pause (parroting)
 - Teilnehmende probieren das Gehörte wiederzugeben
 - Jede modifizierte Datei wird bearbeitet und besprochen
 - Dauer von 75 min
- Vor dem 2. Trainingstermin werden Grafiken für die Teilnehmenden erstellt

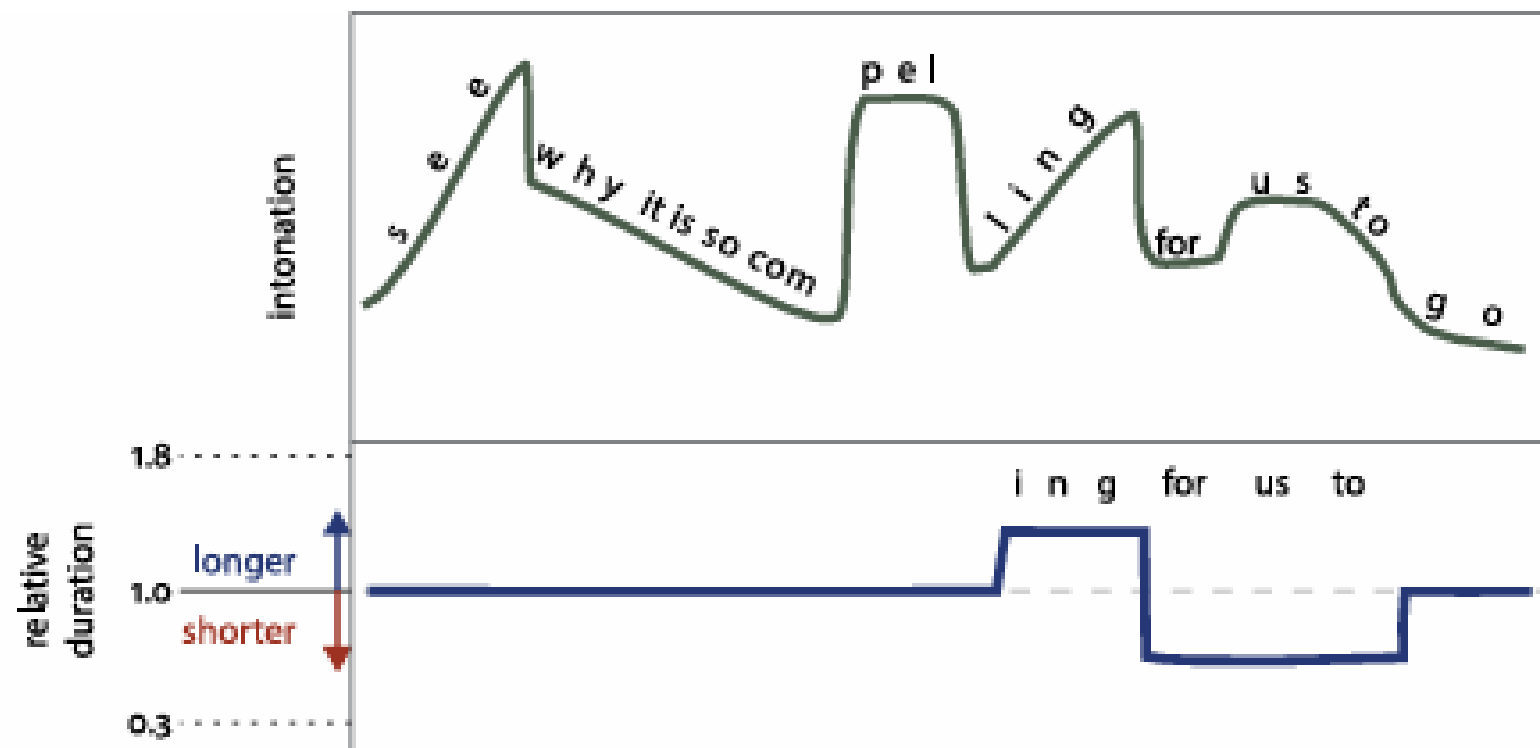


Figure 4. An example of the visual display of one stimulus (based on the recording of speaker F7)

Training

- 2. Trainingstermin:
 - Die Grafiken werden den Teilnehmenden erklärt
 - Die Teilnehmenden sollen die Grafiken „lesen“
 - PSOLA-manipulierte Audiodatei wird vorgespielt
 - 20 sek Pause (parroting)
 - Teilnehmende probieren das Gehörte wiederzugeben
 - Jede modifizierte Datei wird bearbeitet und besprochen
 - Teilnehmende lesen den gesamten Text erneut vor
 - 50 min Dauer
- Einen Monat später lesen die Teilnehmende einen neuen Text vor

Training

- 12 Teilnehmende (8 Frauen, 4 Männer)
- Text 1 vor dem Training
- Text 1 nach dem Training
- Text 2 einen Monat nach Abschluss des Trainings

Wahrgenommene Kompetenz

- Es soll überprüft werden, ob die Teilnehmenden kompetenter wahrgenommen werden
- 3 Texte von 12 Sprechenden = 36 | 45 ähnliche Texte von anderen Sprechenden
- 39 Englischstudierende im ersten Semester bewerteten die Audiodateien auf einer Fünf Punkte Skala

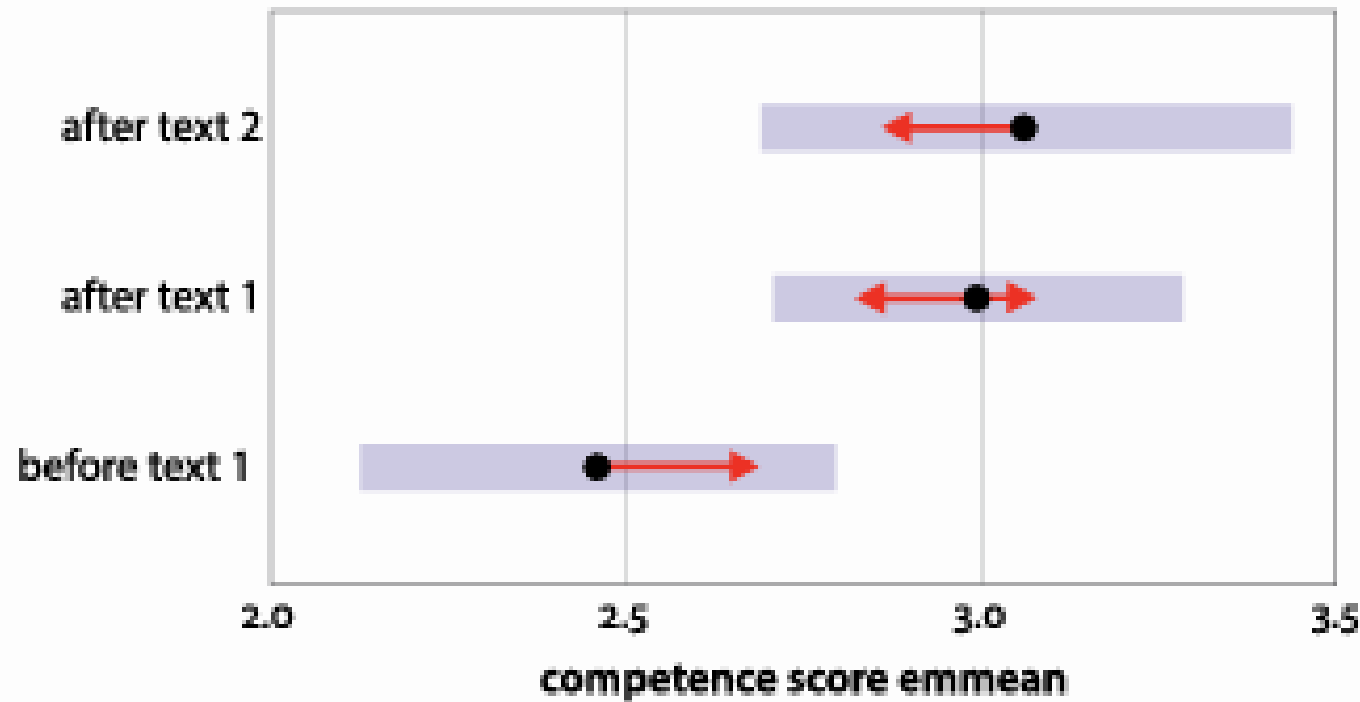


Figure 5. Estimated marginal means of perceived competence scores illustrating the effect of condition. Bars correspond to confidence intervals of the emmeans, and arrows serve for comparisons among them

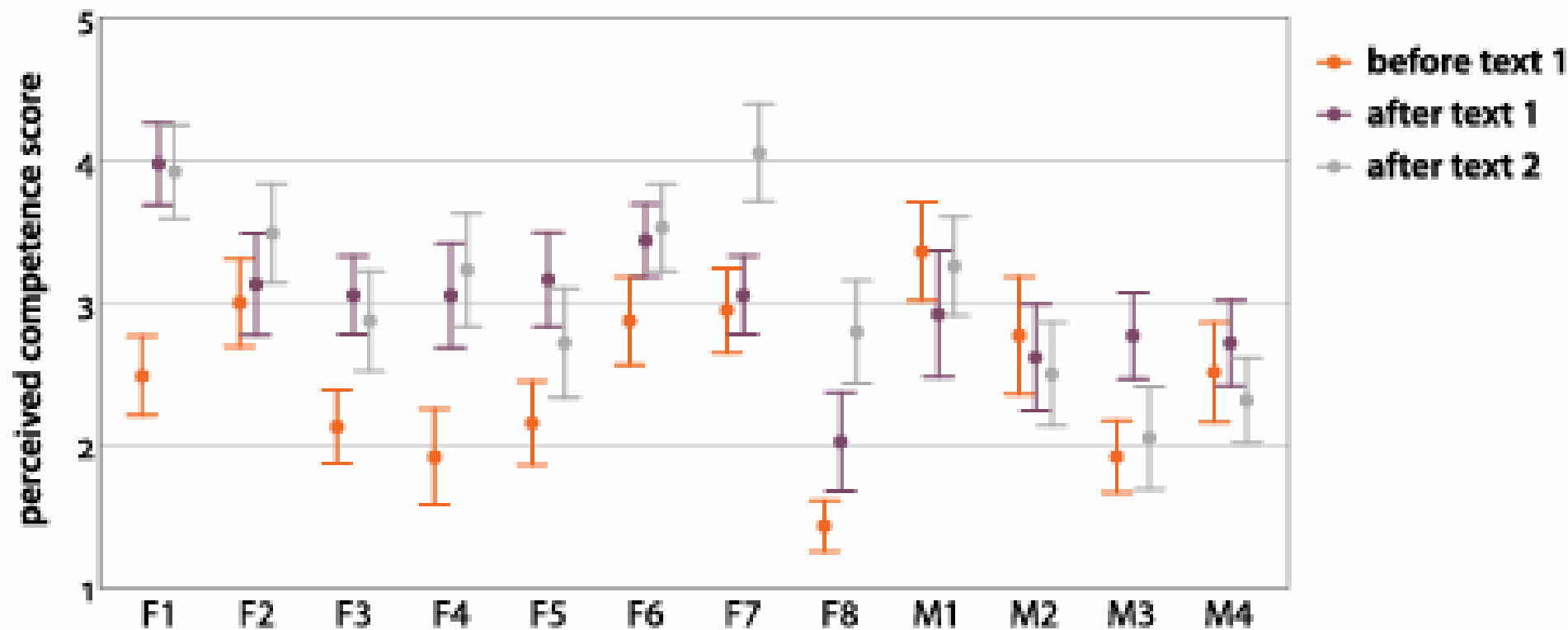


Figure 6. Perceived competence score of individual speakers in three conditions

Wahrgenommene Kompetenz

- Nach dem Training werden die Sprechenden **kompetenter wahrgenommen**
- Der Unterschied zwischen den Texten nach dem Training ist nicht signifikant
- Es bestehen **Unterschiede zwischen den Sprechenden**

Phrasing Analyse

- Es wird angeschaut wie das Gesprochene in Phrasen unterteilt wird
- Bsp: “my answer must at once be, it is no use”
 - Pause nach „be“ notwendig
 - Pause nach „answer“ möglich (je nach Kontext)
 - Pause nach „at“ ist untypisch
- Anhand der Phrasing-Informationen wird versucht, die **Kompetenz der Sprechenden zu schätzen**

Phrasing Analyse

- Die optimale Platzierung prosodischer Grenzen wird von zwei erfahrenen Englischlehrern bestimmt
- In den Tonaufnahmen wird darauf basierend ein *phrasing score* vergeben
- **Je mehr** untypische oder unmögliche prosodische Grenzen, **desto höher** der Score

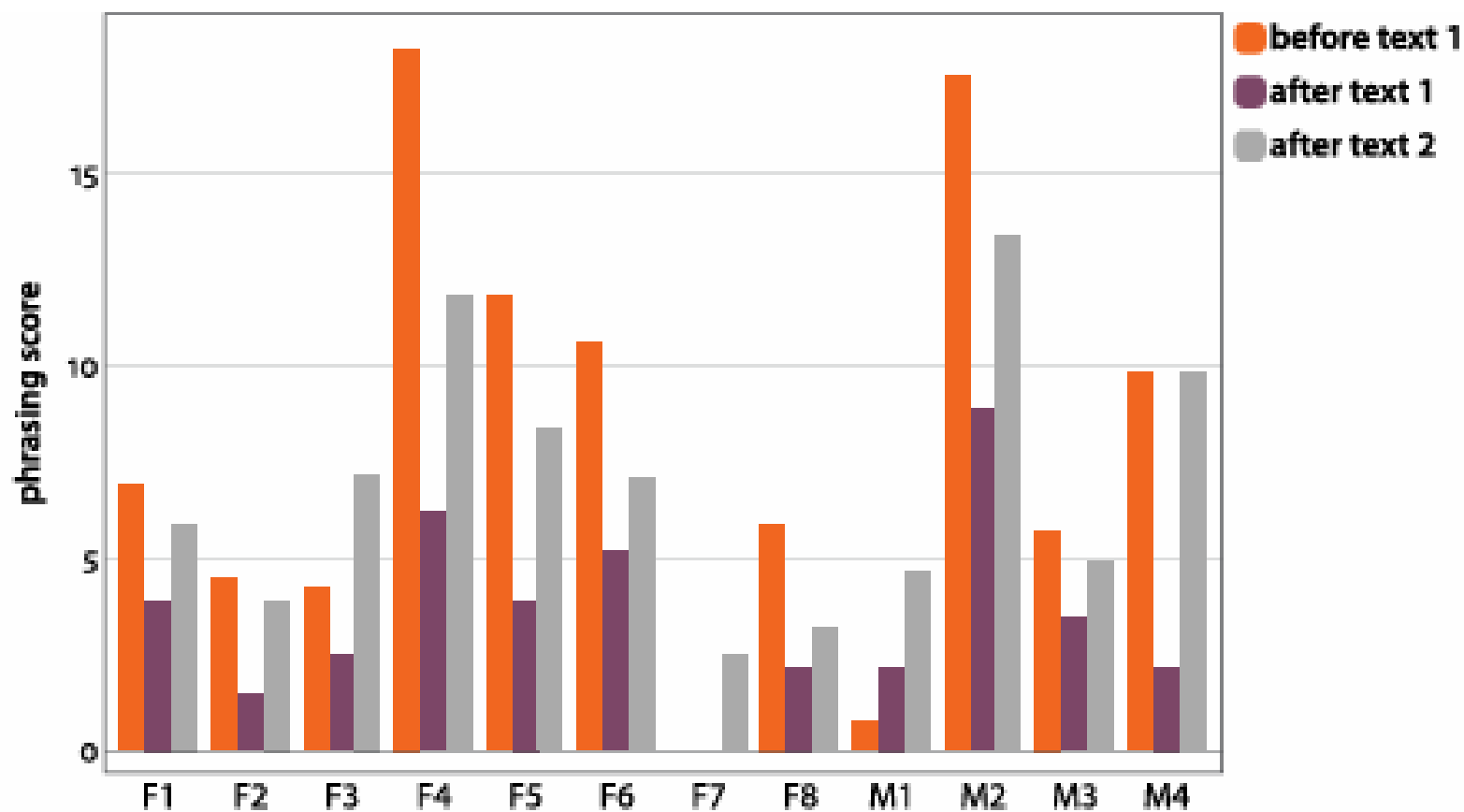


Figure 8. Phrasing score of individual speakers in three conditions

Phrasing Analyse

- Kein Zusammenhang zwischen *phrasing score* und wahrgenommener Kompetenz
- Der *phrasing score* setzt sich aus einem kleinen Teil der Prosodie der Sprechenden zusammen

Produktionsanalyse

- Es wird die **intonatorische Variabilität** angeschaut (wie stark und vielfältig die Tonhöhenbewegung ist)
- Es wird angenommen, dass sich nach dem Training eine höhere Variabilität zeigt

Produktionsanalyse

- Die **Spannweite der Grundfrequenz (f_0)** und der **CSI (*cumulative slope index*)** werden berechnet
- CSI gibt den f_0 -Verlauf über die Zeit an
 - Hoher CSI: viele/starke Tonhöhenbewegungen (f_0 variiert)
 - Niedriger CSI: wenige/schwache Tonhöhenbewegungen (f_0 gleichbleibend)

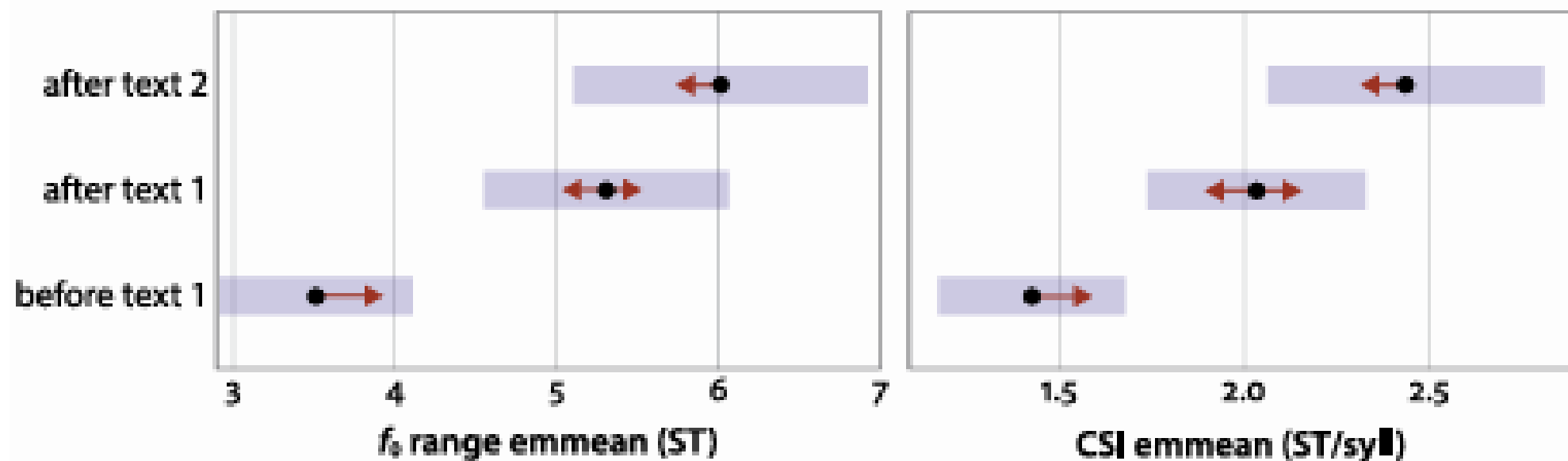


Figure 9. Estimated marginal means of f_0 range (left) and CSI (right) illustrating the effect of condition. Bars correspond to confidence intervals of the emmeans, and arrows serve for comparisons among them

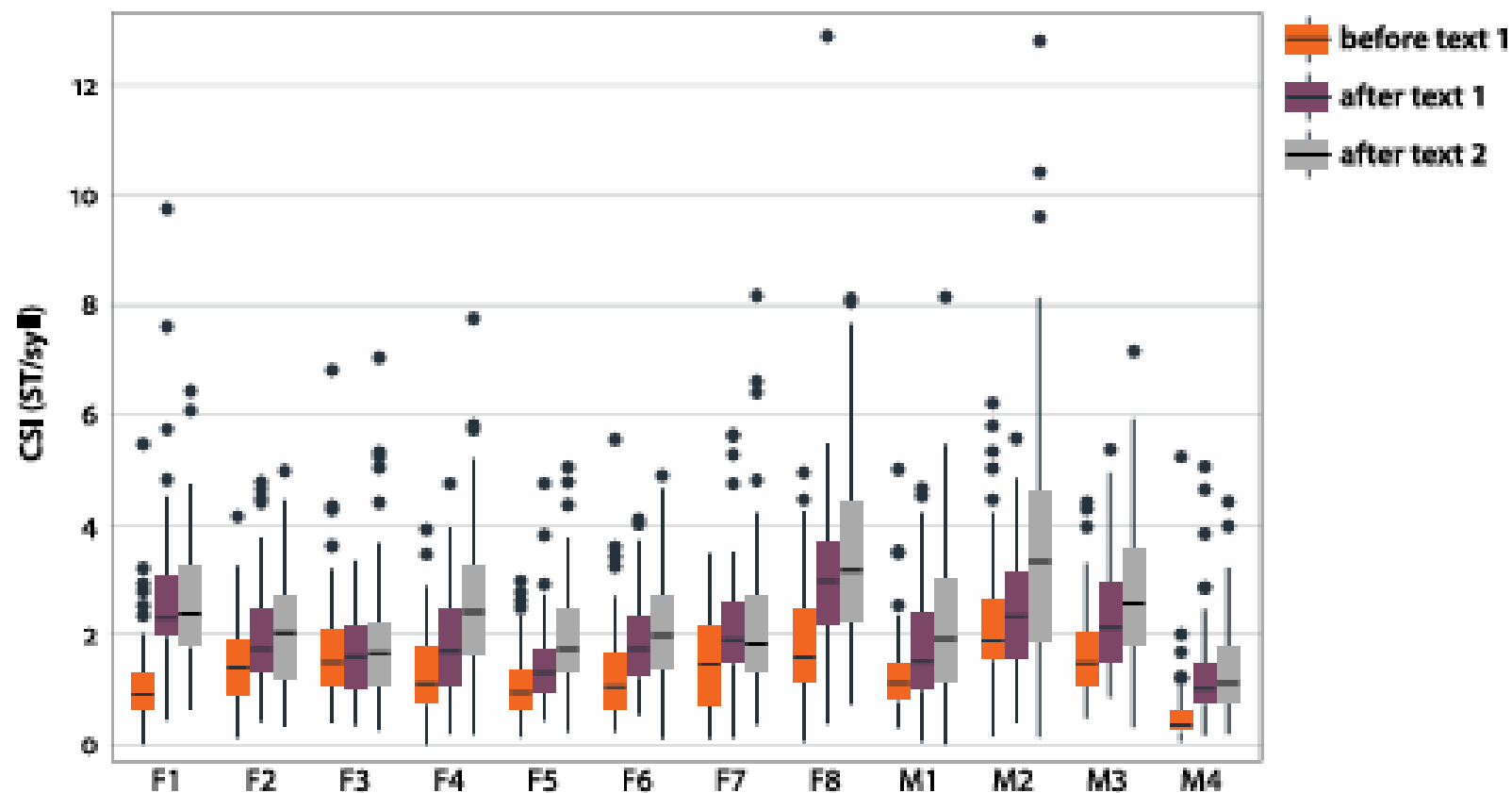


Figure 10. Cumulative slope index (CSI) of individual speakers in three conditions

Produktionsanalyse

- Sowohl die f0-Spannweite als auch der CSI sind nach dem Training signifikant höher

Zusammenfassend

- Sprechende werden nach dem Training kompetenter wahrgenommen
- Die Phrasing Analyse reicht nicht aus, um die Kompetenz der Sprechenden vorherzusagen
- Tonhöhenbewegung ist nach dem Training vielfältiger und stärker

Diskussion

- Training kann sinnvoll sein, um **Sprachunterricht zu ergänzen**, Lehrende müssen ggf. trainiert werden
- Es gab keine Kontrollgruppe
- Es wurde sich nur Vorlesen angeschaut, kein freies Sprechen