The Nature of Speech Motor Representations: Evidence from the Transfer of Learning

Hung-Shao Cheng & Adam Buchwald

New York University

June 21, 2019

Introduction

- **What is the nature of speech motor representations?**
  - When learning novel speech motor patterns, what is being learned?
  - To what extent does the learning **transfer** to other similar but untrained patterns?
- Non-native onset consonant clusters (e.g., GDEEMOO, KTEEMOO)
  - GDEEMOO $\xrightarrow{\text{generalizes}}$ GDAHBEE
  - **Learning is generally at the level of cluster** (Buchwald et al., 2019; Segawa et al., 2019).
- At the level of cluster, what exactly is being learned?
  - General coordination pattern (e.g., stop-stop cluster)?
  - Specific coordination pattern (e.g., GD vs KT)?

Introduction

- **What is the nature of speech motor representations?**
  - When learning novel speech motor patterns, what is being learned?
  - To what extent does the learning **transfer** to other similar but untrained patterns?
- Non-native onset consonant clusters (e.g., **GD**EEMOO, **KT**EEMOO)
  - **GD**EEMOO $\xrightarrow{\text{generalizes}}$ **GD**AHBEE
  - **Learning is generally at the level of cluster** (Buchwald et al., 2019; Segawa et al., 2019).
- At the level of cluster, what exactly is being learned?
  - General coordination pattern (e.g., stop-stop cluster)?
  - Specific coordination pattern (e.g., GD vs KT)?

Introduction

- **What is the nature of speech motor representations?**
  - When learning novel speech motor patterns, what is being learned?
  - To what extent does the learning **transfer** to other similar but untrained patterns?
- Non-native onset consonant clusters (e.g., **GD**EEMOO, **KT**EEMOO)
  - **GD**EEMOO $\xrightarrow{\text{generalizes}}$ **GD**AHBEE
  - **Learning is generally at the level of cluster** (Buchwald et al., 2019; Segawa et al., 2019).
- At the level of cluster, what exactly is being learned?
  - General coordination pattern (e.g., stop-stop cluster)?
  - Specific coordination pattern (e.g., GD vs KT)?

Introduction

- **What is the nature of speech motor representations?**
  - When learning novel speech motor patterns, what is being learned?
  - To what extent does the learning **transfer** to other similar but untrained patterns?
- Non-native onset consonant clusters (e.g., **GD**EEMOO, **KT**EEMOO)
  - **GD**EEMOO $\xrightarrow{\text{generalizes}}$ **GD**AHBEE
  - **Learning is generally at the level of cluster** (Buchwald et al., 2019; Segawa et al., 2019).
- At the level of cluster, what exactly is being learned?
  - General coordination pattern (e.g., stop-stop cluster)?
  - Specific coordination pattern (e.g., GD vs KT)?

Introduction

- **What is the nature of speech motor representations?**
  - When learning novel speech motor patterns, what is being learned?
  - To what extent does the learning **transfer** to other similar but untrained patterns?
- Non-native onset consonant clusters (e.g., **GD**EEMOO, **KT**EEMOO)
  - **GD**EEMOO $\xrightarrow{\text{generalizes}}$ **GD**AHBEE
  - **Learning is generally at the level of cluster** (Buchwald et al., 2019; Segawa et al., 2019).
- At the level of cluster, what exactly is being learned?
  - General coordination pattern (e.g., stop-stop cluster)?
  - Specific coordination pattern (e.g., GD vs KT)?

How specific are speech motor representations?

- Manner of articulation is encoded in speech motor representations.
  - Learning of stops $\not\Rightarrow$ fricatives (Ballard et al., 2007)
- Does training on [voiced, voiceless] clusters transfer to untrained clusters with different voicing?

1. General stop-stop coordination:
   - GD $\Rightarrow$ KT, KT $\Rightarrow$ GD } = **Bi-directional transfer**
   - Same representation (stop-stop clusters)
2. Specific stop-stop coordination:
   - GD $\not\Rightarrow$ KT, KT $\not\Rightarrow$ GD } = **No transfer**
   - Different representation (GD vs KT)

How specific are speech motor representations?

- Manner of articulation is encoded in speech motor representations.
  - Learning of stops $\not\Rightarrow$ fricatives (Ballard et al., 2007)
- Does training on [voiced, voiceless] clusters transfer to untrained clusters with different voicing?

1. General stop-stop coordination:
   - GD $\Rightarrow$ KT, KT $\Rightarrow$ GD } = **Bi-directional transfer**
   - Same representation (stop-stop clusters)
2. Specific stop-stop coordination:
   - GD $\not\Rightarrow$ KT, KT $\not\Rightarrow$ GD } = **No transfer**
   - Different representation (GD vs KT)

How specific are speech motor representations?

- Manner of articulation is encoded in speech motor representations.
  - Learning of stops $\not\Rightarrow$ fricatives (Ballard et al., 2007)
- Does training on [**voiced**, **voiceless**] clusters transfer to untrained clusters with different voicing?

1. General stop-stop coordination:
   - GD $\Rightarrow$ KT, KT $\Rightarrow$ GD $\}$ = **Bi-directional transfer**
   - Same representation (stop-stop clusters)
2. Specific stop-stop coordination:
   - GD $\not\Rightarrow$ KT, KT $\not\Rightarrow$ GD $\}$ = **No transfer**
   - Different representation (GD vs KT)

How specific are speech motor representations?

- Manner of articulation is encoded in speech motor representations.
  - Learning of stops $\not\Rightarrow$ fricatives (Ballard et al., 2007)
- Does training on [**voiced**, **voiceless**] clusters transfer to untrained clusters with different voicing?

1. General stop-stop coordination:
   - $\left.\begin{array}{l} \text{GD} \Rightarrow \text{KT} \\ \text{KT} \Rightarrow \text{GD} \end{array}\right\} =$ **Bi-directional transfer**
   - Same representation (stop-stop clusters)

2. Specific stop-stop coordination:
   - $\left.\begin{array}{l} \text{GD} \not\Rightarrow \text{KT} \\ \text{KT} \not\Rightarrow \text{GD} \end{array}\right\} =$ **No transfer**
   - Different representation (GD vs KT)

How specific are speech motor representations?

- Manner of articulation is encoded in speech motor representations.
  - Learning of stops $\not\Rightarrow$ fricatives (Ballard et al., 2007)
- Does training on [**voiced**, **voiceless**] clusters transfer to untrained clusters with different voicing?

1. General stop-stop coordination:
   - $\left.\begin{array}{l} \text{GD} \Rightarrow \text{KT} \\ \text{KT} \Rightarrow \text{GD} \end{array}\right\} =$ **Bi-directional transfer**
   - Same representation (stop-stop clusters)

2. Specific stop-stop coordination:
   - $\left.\begin{array}{l} \text{GD} \not\Rightarrow \text{KT} \\ \text{KT} \not\Rightarrow \text{GD} \end{array}\right\} =$ **No transfer**
   - Different representation (GD vs KT)

Does complexity affect transfer?

- Voiced clusters are produced with lower accuracy rate (Davidson, 2006) and hypothesized to be more complex:
  - Phonologically
    - Typologically *marked* (Morelli, 1999)
  - Phonetically:
    - Aerodynamically more challenging (Ohala, 1997)
      - Additional level of coordination
- If complexity matters:
  - $\left.\begin{array}{l} GD \Rightarrow KT \\ KT \not\Rightarrow GD \end{array}\right\} =$ **Uni-directional transfer**

Does complexity affect transfer?

- Voiced clusters are produced with lower accuracy rate (Davidson, 2006) and hypothesized to be more complex:
  - Phonologically
    - Typologically *marked* (Morelli, 1999)
  - Phonetically:
    - Aerodynamically more challenging (Ohala, 1997)
      - Additional level of coordination
- If complexity matters:
  - $\left.\begin{array}{l} GD \Rightarrow KT \\ KT \not\Rightarrow GD \end{array}\right\} =$ **Uni-directional transfer**

Does complexity affect transfer?

- Voiced clusters are produced with lower accuracy rate (Davidson, 2006) and hypothesized to be more complex:
  - Phonologically
    - Typologically *marked* (Morelli, 1999)
  - Phonetically:
    - Aerodynamically more challenging (Ohala, 1997)
      - Additional level of coordination
- If complexity matters:
  - $\left.\begin{array}{l} GD \Rightarrow KT \\ KT \not\Rightarrow GD \end{array}\right\} =$ **Uni-directional transfer**

Does complexity affect transfer?

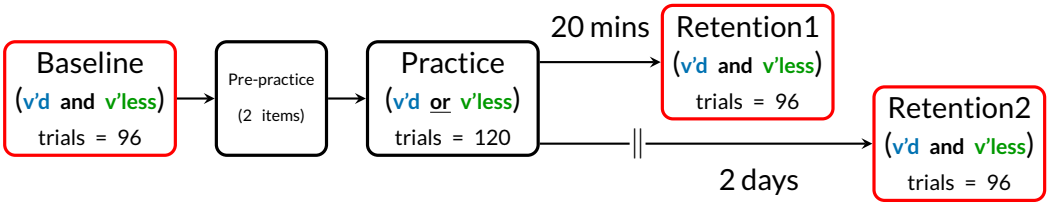
- Voiced clusters are produced with lower accuracy rate (Davidson, 2006) and hypothesized to be more complex:
  - Phonologically
    - Typologically *marked* (Morelli, 1999)
  - Phonetically:
    - Aerodynamically more challenging (Ohala, 1997)
      - Additional level of coordination
- If complexity matters:
  - $\left.\begin{array}{l} \text{GD} \Rightarrow \text{KT} \\ \text{KT} \not\Rightarrow \text{GD} \end{array}\right\} =$ **Uni-directional transfer**

Does complexity affect transfer?

- Voiced clusters are produced with lower accuracy rate (Davidson, 2006) and hypothesized to be more complex:
  - Phonologically
    - Typologically *marked* (Morelli, 1999)
  - Phonetically:
    - Aerodynamically more challenging (Ohala, 1997)
      - Additional level of coordination
- If complexity matters:
  - $\left.\begin{array}{l} \text{GD} \Rightarrow \text{KT} \\ \text{KT} \not\Rightarrow \text{GD} \end{array}\right\} =$ **Uni-directional transfer**

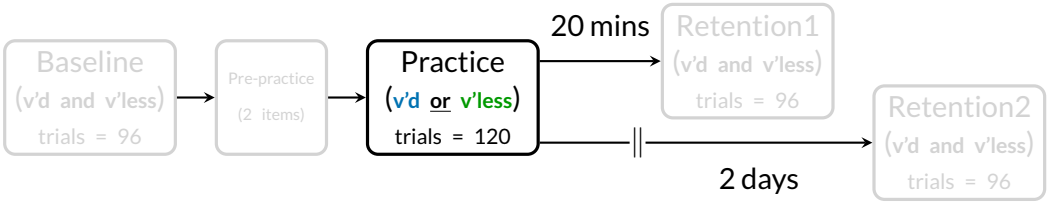
Paradigm

- Practice-based nonword production with orthographic and auditory models.
- Random & variable practice (Maas et al., 2008) without feedback:
  - Voiced condition (/db/, /gb/, /gd/; n = 10)
  - Voiceless condition (/tp/, /kp/, /kt/; n = 10)
- Probed on all voiced and voiceless clusters at the baseline, retention 1, retention 2
- Pre-practice: Information about how to produce non-native clusters



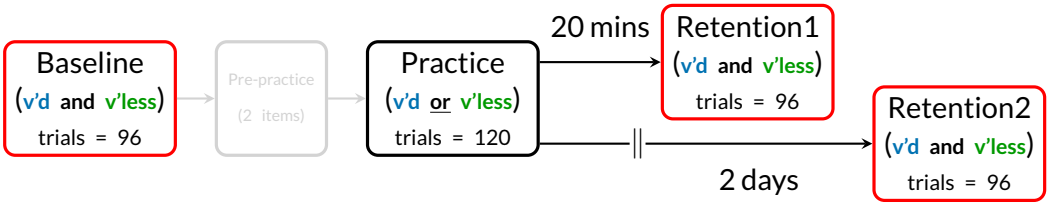
Paradigm

- Practice-based nonword production with orthographic and auditory models.
- Random & variable practice (Maas et al., 2008) without feedback:
  - Voiced condition (/db/, /gb/, /gd/; n = 10)
  - Voiceless condition (/tp/, /kp/, /kt/; n = 10)
- Probed on all voiced and voiceless clusters at the baseline, retention 1, retention 2
- Pre-practice: Information about how to produce non-native clusters



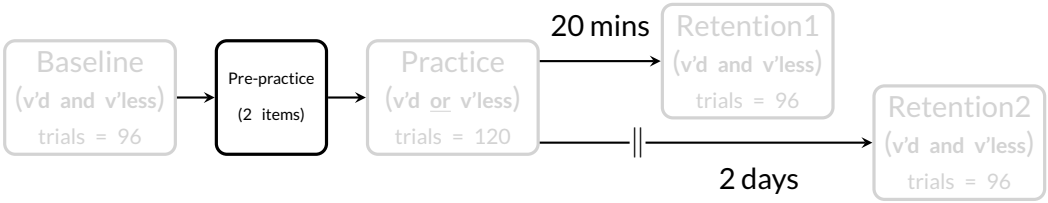
Paradigm

- Practice-based nonword production with orthographic and auditory models.
- Random & variable practice (Maas et al., 2008) without feedback:
  - Voiced condition (/db/, /gb/, /gd/; n = 10)
  - Voiceless condition (/tp/, /kp/, /kt/; n = 10)
- Probed on all voiced and voiceless clusters at the baseline, retention 1, retention 2
- Pre-practice: Information about how to produce non-native clusters



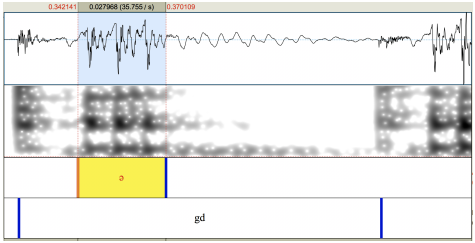
Paradigm

- Practice-based nonword production with orthographic and auditory models.
- Random & variable practice (Maas et al., 2008) without feedback:
  - Voiced condition (/db/, /gb/, /gd/; n = 10)
  - Voiceless condition (/tp/, /kp/, /kt/; n = 10)
- Probed on all voiced and voiceless clusters at the baseline, retention 1, retention 2
- Pre-practice: Information about how to produce non-native clusters



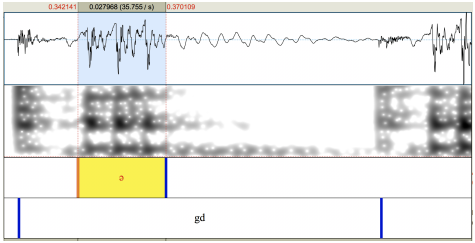
Measurements

- **Accuracy:**
  - Cluster accuracy (agreement: 91%)
  - Acoustically-informed transcription (Wilson et al., 2014)
  - Epenthetic vowel $\equiv$
    1. $\geq$ Two repetitive vocoid cycles.
    2. Higher formant structures (e.g., F2, F3)
  - Coders blinded to condition and session
- **Motor acuity:**
  - Burst-to-burst duration (agreement: 96%):
  - $\downarrow$ duration = better coordination of the cluster



Measurements

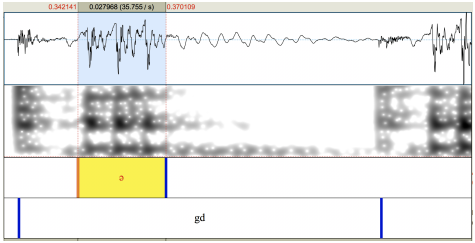
- **Accuracy**:
  - Cluster accuracy (agreement: 91%)
  - Acoustically-informed transcription (Wilson et al., 2014)
  - Epenthetic vowel $\equiv$
    1. $\geq$ Two repetitive vocoid cycles.
    2. Higher formant structures (e.g., F2, F3)
  - Coders blinded to condition and session



- **Motor acuity**:
  - Burst-to-burst duration (agreement: 96%):
  - $\downarrow$ duration = better coordination of the cluster

Measurements

- **Accuracy**:
  - Cluster accuracy (agreement: 91%)
  - Acoustically-informed transcription (Wilson et al., 2014)
  - Epenthetic vowel $\equiv$
    1. $\geq$ Two repetitive vocoid cycles.
    2. Higher formant structures (e.g., F2, F3)
  - Coders blinded to condition and session



- **Motor acuity**:
  - Burst-to-burst duration (agreement: 96%):
  - $\downarrow$ duration = better coordination of the cluster

Statistical approach to detect transfer

- Mixed-effects logistic models for cluster accuracy
  - accuracy $\sim$ **Session*****Type** + Cluster + (1 | Participant) + (1 | Item)
    - Accuracy: correct versus incorrect
    - Session: baseline, retention1, retention2
    - Type: trained versus untrained (transfer)
    - The best fitting model was selected by AIC and **BIC** scores (Harel and McAllister, 2019).
- Transfer (✓) $\equiv$ main effect of Session and no Session by Type interaction

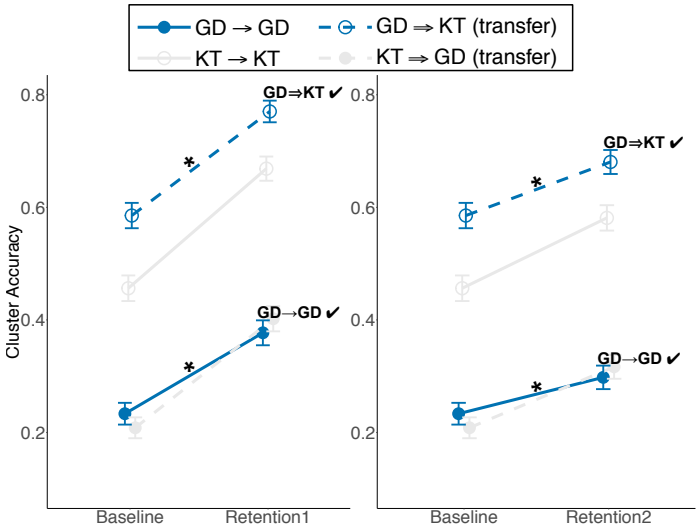
Statistical approach to detect transfer

- Mixed-effects logistic models for cluster accuracy
  - accuracy $\sim$ **Session*Type** + Cluster + (1 | Participant) + (1 | Item)
    - Accuracy: correct versus incorrect
    - Session: baseline, retention1, retention2
    - Type: trained versus untrained (transfer)
    - The best fitting model was selected by AIC and **BIC** scores (Harel and McAllister, 2019).
- Transfer (✓) $\equiv$ main effect of Session and no Session by Type interaction

Statistical approach to detect transfer

- Mixed-effects logistic models for cluster accuracy
  - accuracy $\sim$ **Session*Type** + Cluster + (1 | Participant) + (1 | Item)
    - Accuracy: correct versus incorrect
    - Session: baseline, retention1, retention2
    - Type: trained versus untrained (transfer)
    - The best fitting model was selected by AIC and **BIC** scores (Harel and McAllister, 2019).
- Transfer (✓) $\equiv$ main effect of Session and no Session by Type interaction

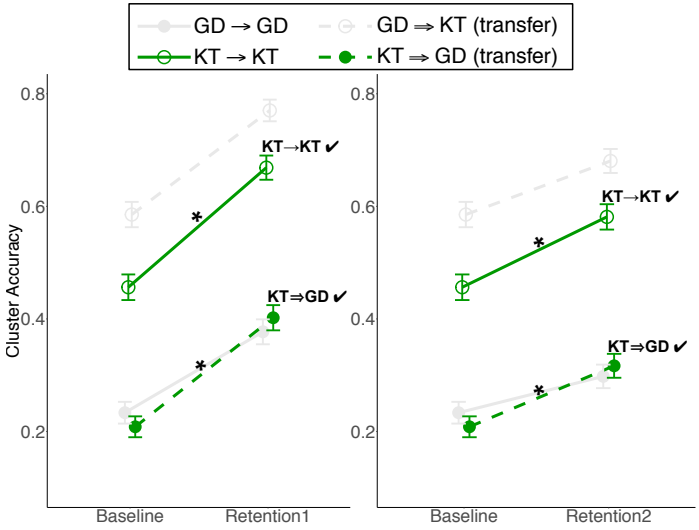
Cluster accuracy



Accuracy ∼ Session*Type + Cluster + (1 | Participant) + (1 | Item)

- Voiced condition: session effect not affected by type at both R1 and R2
  - GD ⇒ KT ✓
- Voiceless condition: session effect not affected by type at both R1 and R2
  - KT ⇒ GD ✓
- Bi-directional transfer

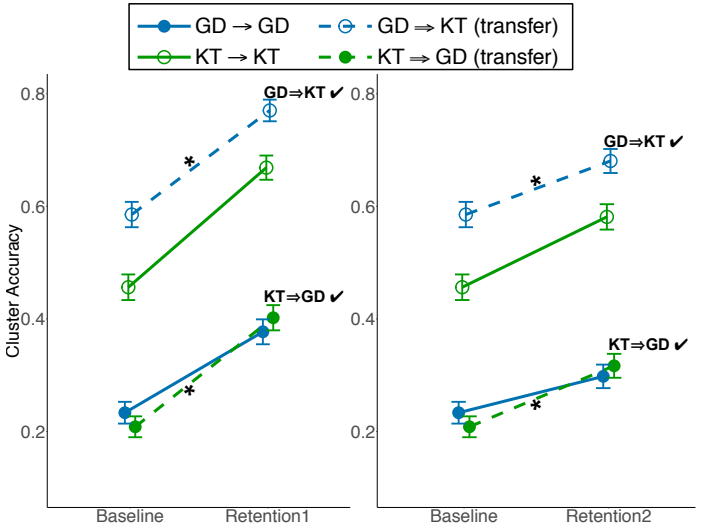
Cluster accuracy



Accuracy ∼ Session*Type + Cluster + (1 | Participant) + (1 | Item)

- Voiced condition: session effect not affected by type at both R1 and R2
  - GD ⇒ KT ✓
- Voiceless condition: session effect not affected by type at both R1 and R2
  - KT ⇒ GD ✓
- Bi-directional transfer

Cluster accuracy

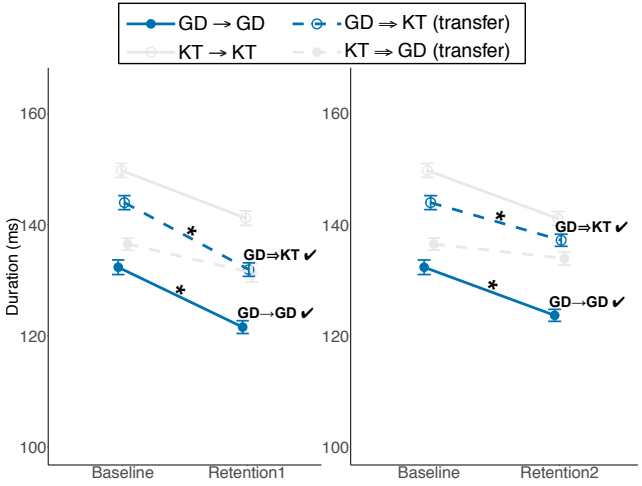


Accuracy ∼ Session*Type + Cluster + (1 | Participant) + (1 | Item)

- Voiced condition: session effect not affected by type at both R1 and R2
  - GD ⇒ KT ✓
- Voiceless condition: session effect not affected by type at both R1 and R2
  - KT ⇒ GD ✓
- Bi-directional transfer

Burst-to-burst duration

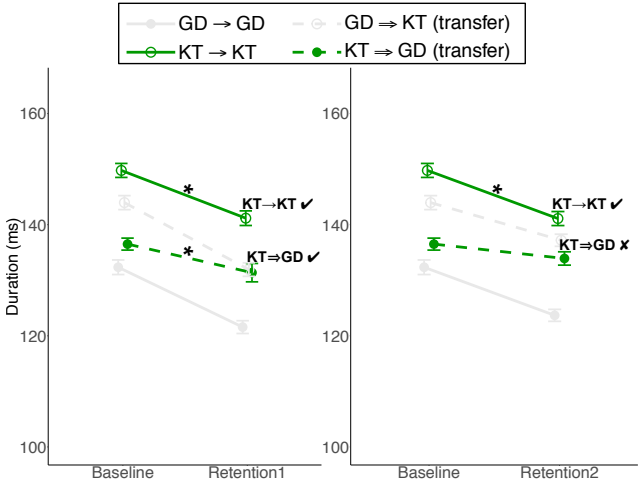
Burst-to-burst duration



- Voiced condition: session effect not affected by type at both R1 and R2
  - GD ⇒ KT ✓
- Voiceless condition: session effect affected by type at both R1 and R2
  - KT ⇒ GD ✗
- No significant Condition by Type by Session interaction
  - no difference in transfer magnitude between conditions

Duration ∼ Session*Type + Cluster Accuracy + Cluster + (1 | Participant) + (1 | Item)

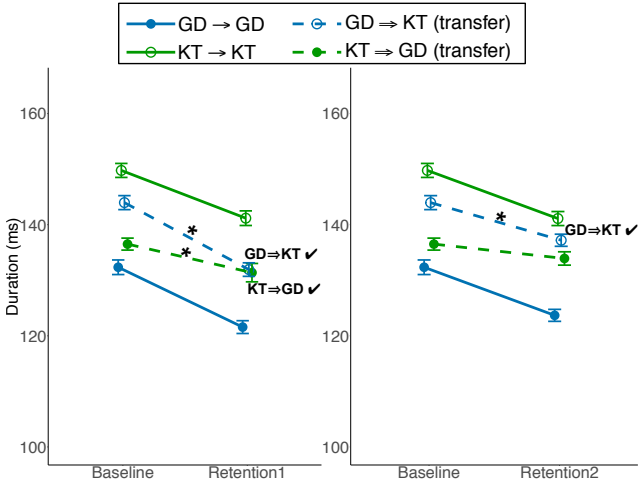
Burst-to-burst duration



- Voiced condition: session effect not affected by type at both R1 and R2
  - GD ⇒ KT ✓
- Voiceless condition: session effect affected by type at both R1 and R2
  - KT ⇒ GD ✗
- No significant Condition by Type by Session interaction
  - no difference in transfer magnitude between conditions

Duration ∼ Session*Type + Cluster Accuracy + Cluster + (1 | Participant) + (1 | Item)

Burst-to-burst duration



- Voiced condition: session effect not affected by type at both R1 and R2
  - GD ⇒ KT ✓
- Voiceless condition: session effect affected by type at both R1 and R2
  - KT ⇒ GD ✗
- **No significant Condition by Type by Session interaction**
  - no difference in transfer magnitude between conditions

Duration ∼ Session*Type + Cluster Accuracy + Cluster + (1 | Participant) + (1 | Item)

Discussion

- Measureable learning gains on trained clusters
- Bi-directional transfer across voicing categories
  - GD $\Rightarrow$ KT
    KT $\Rightarrow$ GD
- **General coordination of stop-stop cluster is encoded in the speech motor representations**

Discussion

- Measureable learning gains on trained clusters
- Bi-directional transfer across voicing categories
  - GD $\Rightarrow$ KT
    KT $\Rightarrow$ GD
- General coordination of stop-stop cluster is encoded in the speech motor representations

Discussion

- Measureable learning gains on trained clusters
- Bi-directional transfer across voicing categories
  - GD $\Rightarrow$ KT
    KT $\Rightarrow$ GD
- General coordination of stop-stop cluster is encoded in the speech motor representations

Discussion

- Measureable learning gains on trained clusters
- Bi-directional transfer across voicing categories
  - GD $\Rightarrow$ KT
    KT $\Rightarrow$ GD
- **General coordination of stop-stop cluster is encoded in the speech motor representations**

Acknowledgement

- We would like to thank
  - **Dr. Tara McAllister**
  - **Members of Speech Reading Group**
  - **Members of PLAN lab**
  - **Alexandra Gordon, Izabela Grzebyk, & Kevin Tjokro**

Acknowledgement

Questions?

References

Ballard, K. J., Maas, E., and Robin, D. A. (2007). Treating control of voicing in apraxia of speech with variable practice. *Aphasiology*, 21(12):1195–1217.

Buchwald, A., Calhoun, H., Rimikis, S., Lowe, M. S., Wellner, R., and Edwards, D. J. (2019). Using tdcs to facilitate motor learning in speech production: The role of timing. *Cortex*, 111:274–285.

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, 34(1):104–137.

Harel, D. and McAllister, T. (2019). Multilevel models for communication sciences and disorders. *Journal of Speech, Language, and Hearing Research*, pages 1–19.

Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., and Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*.

Morelli, F. (1999). *The phonotactics and phonology of obstruent clusters in Optimality Theory*. PhD thesis.

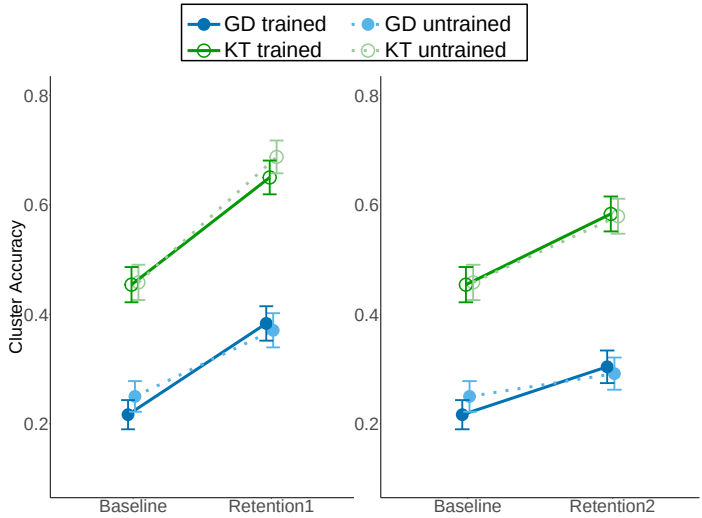
References

Ohala, J. J. (1997). Aerodynamics of phonology. In *Proceedings of the Seoul International Conference on Linguistics*, volume 92, page 97.

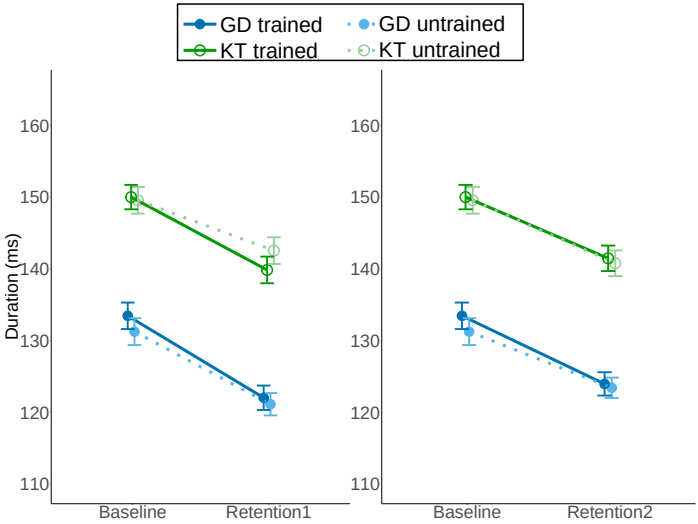
Segawa, J., Masapollo, M., Tong, M., Smith, D. J., and Guenther, F. H. (2019). Chunking of phonological units in speech sequencing. *Brain and Language*, 195:104636.

Wilson, C., Davidson, L., and Martin, S. (2014). Effects of acoustic–phonetic detail on cross-language speech production. *Journal of Memory and Language*, 77:1–24.

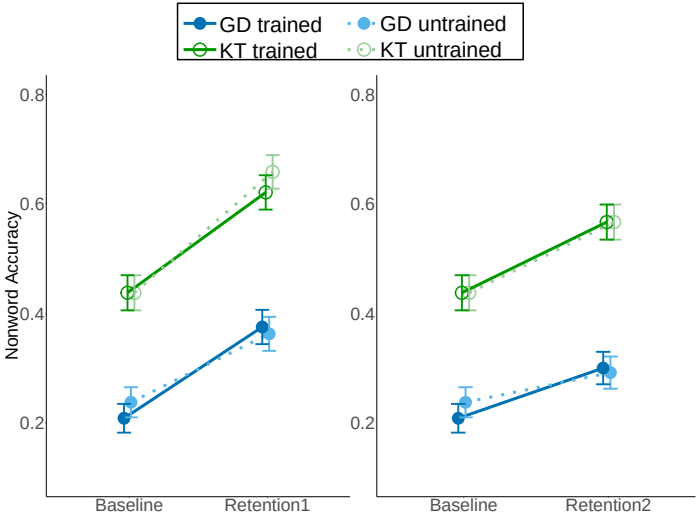
Cluster accuracy: generalization



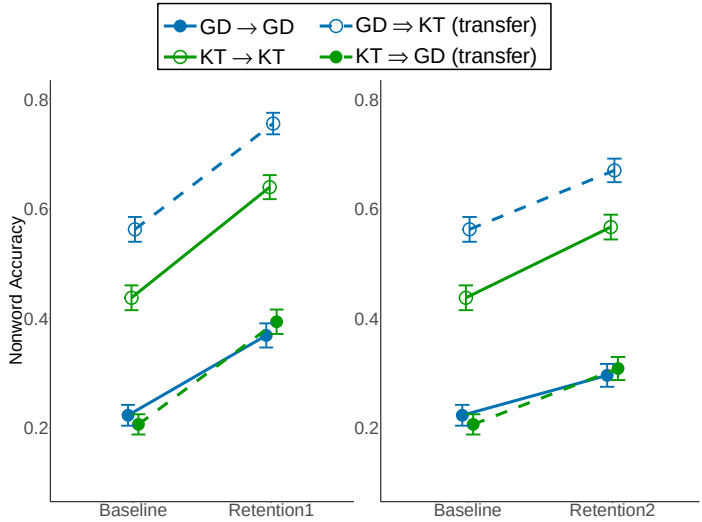
Burst to burst duration: generalization



Nonword accuracy: generalization

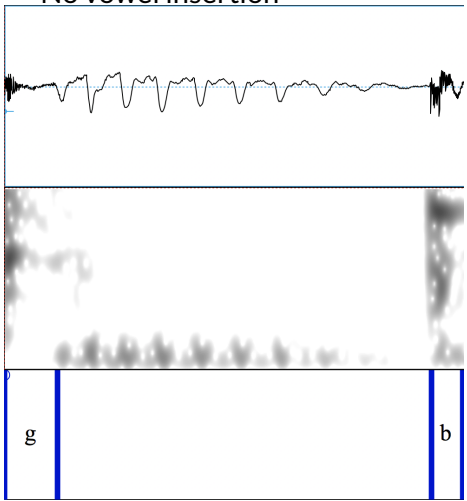


Nonword accuracy: transfer of learning

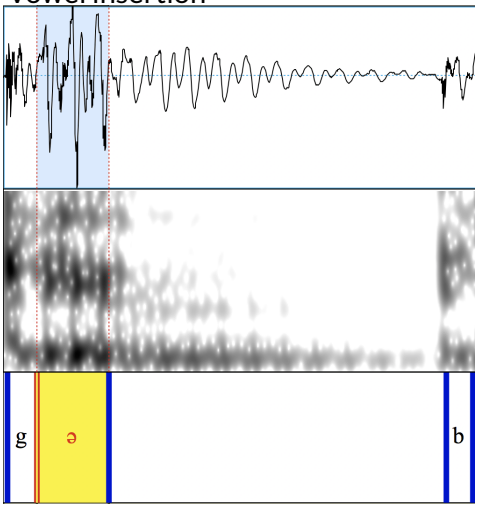


Vowel epenthesis example

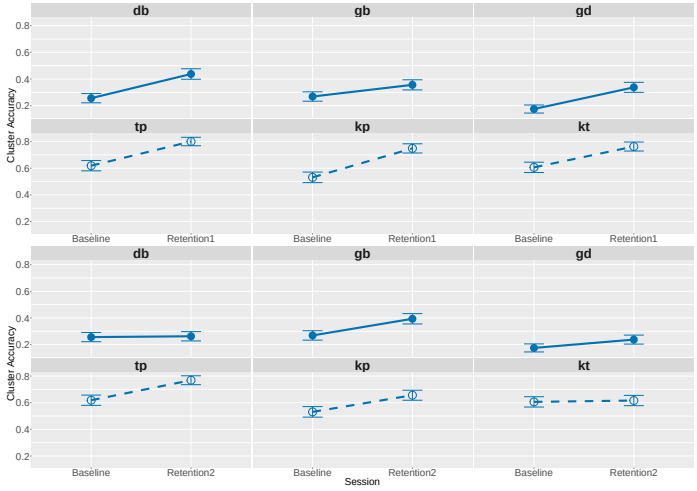
- No vowel insertion



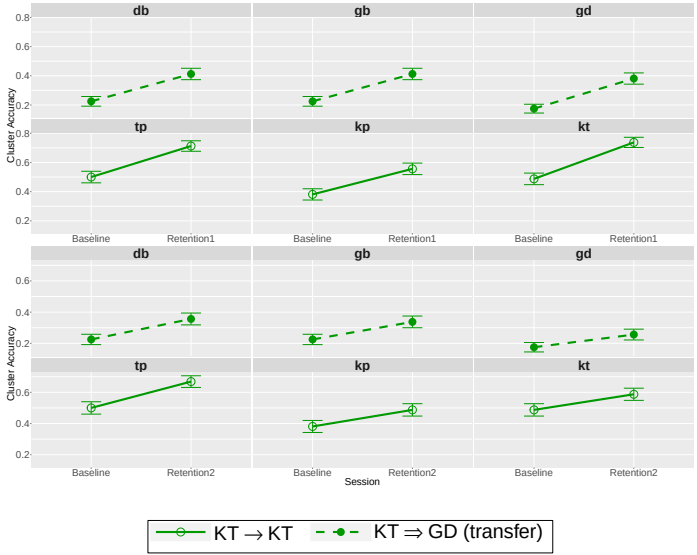
- Vowel insertion



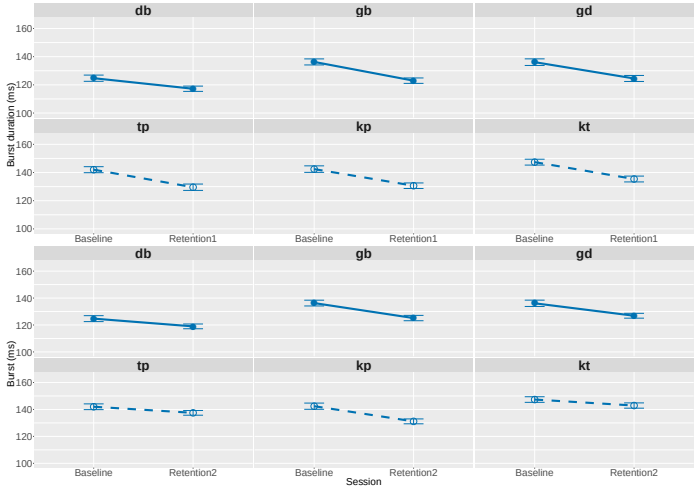
Voiced condition: cluster accuracy by types



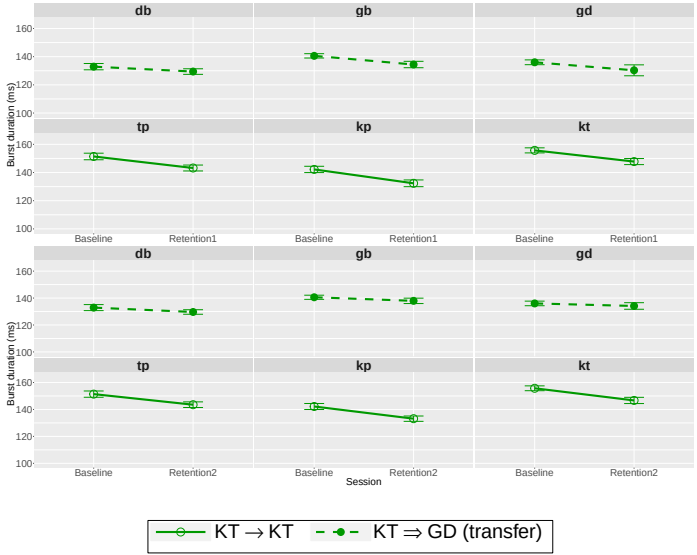
Voiceless condition: cluster accuracy by types



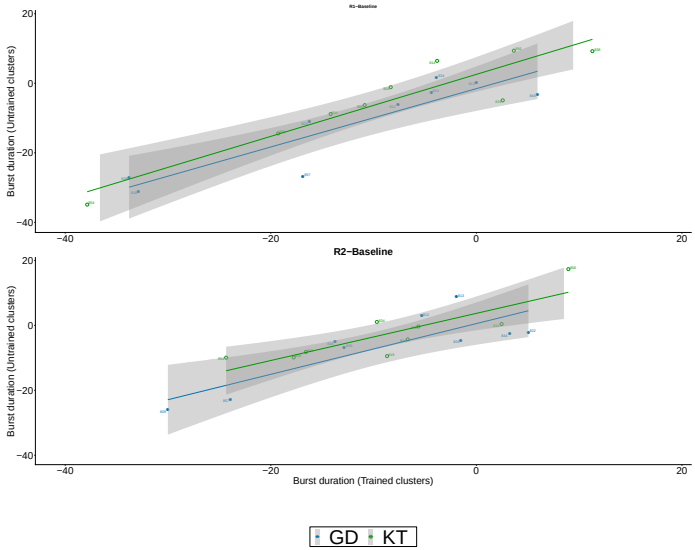
Voiced condition: burst duration by types



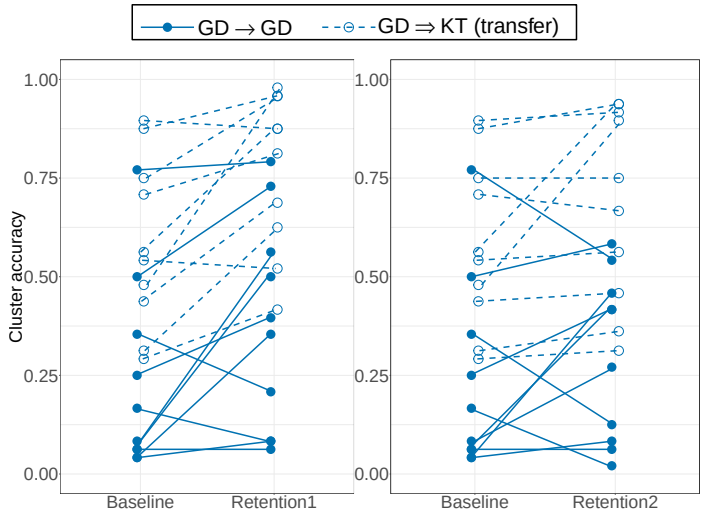
Voiceless condition: burst to burst duration by types



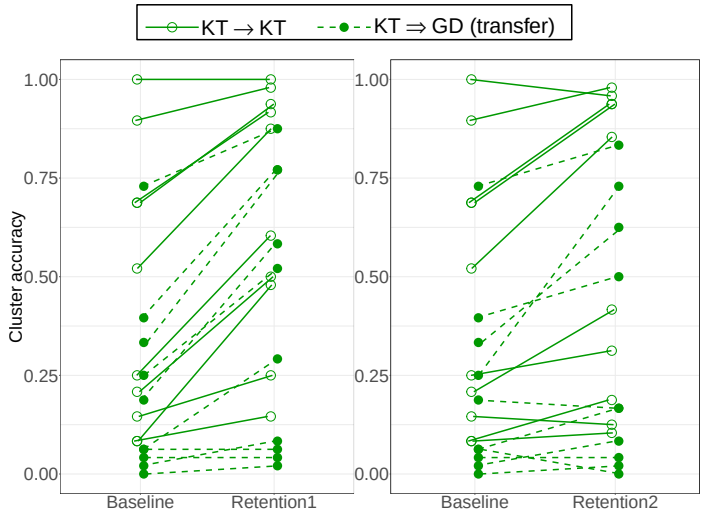
Magnitude of transfer: burst to burst duration



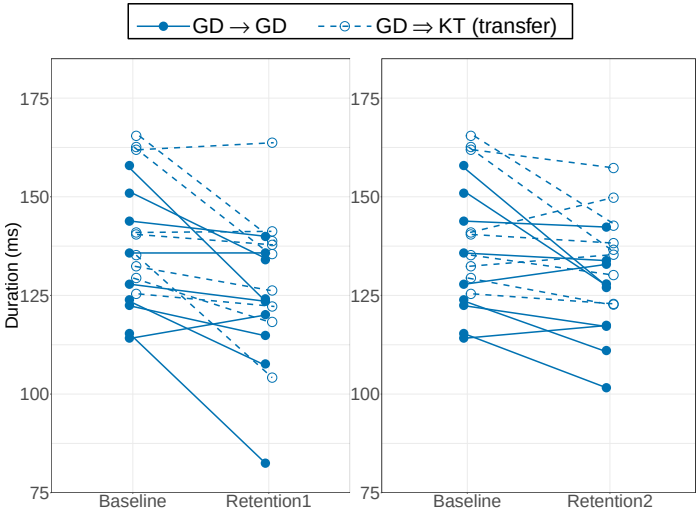
Cluster accuracy: Voiced condition individual data



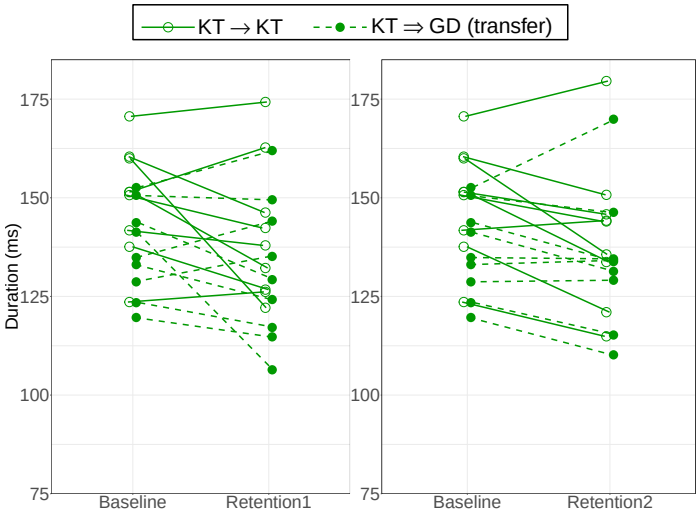
Cluster accuracy: Voiceless condition individual data



Burst to Burst duration: Voiced condition individual data



Burst to Burst duration: Voiceless condition individual data



Numbers of tokens for burst to burst duration

		Correct		Epenthetic vowel	
		Trained	Untrained	Trained	Untrained
Voiced condition	Baseline	94	245	314	165
	Retention1	172	331	226	83
	Retention2	138	293	282	130
Voiceless condition	Baseline	179	80	250	340
	Retention1	271	157	150	259
	Retention2	232	130	196	298