Syllable-pointing gesture coordination in Polish counting out rhymes: The effect of speech rate

Katarzyna Stoltmann & Susanne Fuchs

Leibniz-Centre General Linguistics (ZAS)
Schützenstrasse 18, 10117 Berlin, Germany

stoltmann@leibniz-zas.de; fuchs@leibniz-zas.de

Abstract

We investigated the stability of the relationship between number of syllables and pointing gestures under different speech rate constraints. Participants of the study realized Polish counting out rhymes at normal and fast speech rates. Pointing gestures of the apex finger were recorded with a motion capture system and speech acoustics were recorded simultaneously. We hypothesized that for stable coordination of syllable realization and pointing gestures, either movement amplitude would be smaller or peak velocity would be higher at a faster rate. Unstable coordination due to temporal constraints could lead to a coordinational reorganization of speech and finger motions. Results indicate a large degree of stability, which is, however, speaker-specific. Most speakers realize comparable amounts of syllables per pointing gesture under both speech rate conditions. They frequently shorten duration and increase the speed of the pointing gesture.

1 Introduction

Our work is concerned with a deeper understanding of speech production in the context of rhythmic motor actions and their mutual influence (for a general review see Iversen & Balasubramaniam, 2016). Speech and motor actions often flexibly coordinate with another. A common principle to test the stability and flexibility of coordinated behaviour is to apply constraints that may perturb or challenge these actions. For example, Rochet-Capellan & Schwartz (2008) or Lanica & Fuchs (2011) used the speeding paradigm to investigate the coordination between jaw, lower lip and tongue tip movement in CVCV utterances with either a coronal consonant in the first syllable and a bilabial in the second or vice versa. With increase in speed, speakers change from two jaw cycles to one, which in some cases leads to a reorganization of the tongue tip-jaw-lower lip coordination. The reduction in jaw cycles may be the consequence of the dynamic properties of the jaw. It is the heaviest articulator, because it includes bones, while the tongue tip and the lower lip consist of light and flexible soft tissue and their motion is particularly fast (Jannedy et al., 2010). Thus, there are some restrictions on the jaw when it comes to oscillating with a high frequency. Similarly, the dynamics of speech and the arm's pointing gesture movement differ, although they belong to the same body. Since the arm is heavier, arm motions are slower than speech movements. This may be one reason why in some coordinated speech-pointing gesture actions (Rochet-Capellan et al., 2008), the arm motion starts earlier than the speech motion and the faster system (speech articulation) adjusts quicklier at places where cordination among the two is needed (e.g. in the vowel of a stressed syllable). It has further been proposed that certain frequency relations between different motor systems are more optimal than others (e.g., Cummins & Simko, 2009). For speech-pointing gestures, Rochet Cappelan et al. (2007) suggested 2:1 as an optimal ratio and provided first evidence for this. In our own work on German counting out rhymes (Fuchs & Reichel, 2016), we were able to furthermore confirm that the relation between the number of syllables and pointing gestures is stable, though to some extent speaker-specific (see Figure 4 in Fuchs & Reichel, 2016). However, only five speakers were recorded.

Moreover, the relation between the number of syllables and pointing gestures may also be driven by the speech material. If words consist only of one syllable, they may well correspond to one pointing gestures (stroke). If words consist of two syllables, either one or two pointing gestures may be realized. For words consisting of 3 syllables, we would expect either 1, 2 or 3 pointing gestures. Counting out rhymes are a fascinating object of investigation in this respect, because they are a natural testbed for studying this relation. They are often constructed in such a sense that the number of syllables per line changes so that it is rather unpredictable who will win the game and who will be out. Although regional differences exist, widespread knowledge of counting out rhymes appears to exist spanning populations of varying social backgrounds. They are learned in early childhood and allow for investigation of the development of speech with pointing gestures throughout the life span. They exist in many cultures and are part of the oral poetry tradition (Hanna et al., 2002). Cross-linguistic comparisons can be made while taking language specific prosody and linguistic structure into account. The language, we will study, Polish, has predictable stress patterns on the penultimate syllable (Malisz et al. 2013).

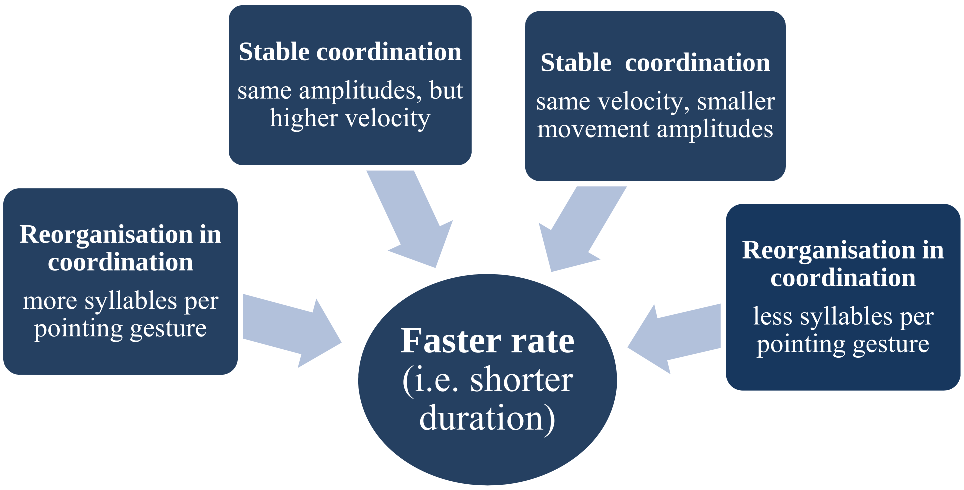

Figure 1. Schematic view of expected speech-pointing gesture relation with increasing speech rate.

Various scenarios for the potential relation between speech and pointing gestures are possible. They are depicted in Figure 1. All of them lead to a faster rate, i.e. a shortening of the temporal window in which the respective counting out rhyme is realized. If the coordination between speech and pointing gestures is stable, we could expect that pointing gestures are reduced in amplitude from one stroke to the next, while peak velocity stays the same. Alternatively, movement amplitudes could also stay the same while peak velocity within a stroke is increased. If the coordination between speech and pointing gesture is more flexible, we expect a reorganization of the two motor actions. Since the speech motor system is a faster oscillating system than the arm motor system, the most economical reorganization would be to produce more syllables per pointing gesture with increasing rate. Another solution could be that speech and hand gestures may reorganize when their relation is not 1:1, i.e. the number of pointing gestures increases so that both systems oscillate with a similar frequency. Such a scenario could be the consequence of a stronger coupling between the two motor systems, similar to the pattern of head motions that has been observed with an increase in speech rate while producing repetitive utterance (Mark Tiede, unpublished data).

2 Methods of data acquisition, annotation and analysis

2.1 Experimental set-up and tasks

Participants stood in front of a chair with a teddy bear and were instructed to play the counting out rhyme game with a teddy bear as a fictitious person. Pointing gestures were measured by means of a motion capture system (OptiTrack, *Motive* Version 1.9.0) with 12 cameras (Prime 13). Motion data was captured at a 200 Hz sampling frequency and simultaeous acoustic data was captured with 44.1 kHz. 14 markers were placed on a frontlet (one anterior, one posterior and one at the right lateral side), one on the tip of each index finger and one at the bottom of each index finger, one at each wrist, one at each elbow, one at each shoulder (on top of the acromioclavicular ligament), one at the height of the C7 vertebra). An additional marker was placed at the nose of the bear (see Figure 2 for a general overview). The participants' task was: a) to read the rhymes aloud without any hand motion; b) to say the rhymes again while pointing with the nondominant (or dominant hand for the first 6 speakers), d) to point with the dominant hand while saying the rhymes (or the non-dominant one for the first 6 speakers) +e) + f) and finally to repeat the previous two tasks with increased speech rate (time constraints).

2.2 Participants, and speech material

Participants were native speakers of Polish who had been residents of Berlin for no longer than 6 months (most of them were Erasmus students). They were recruited via international contact points at different universities in Berlin. Their age ranged from 21-27 (mean 24.1 years) and all were right handed according to the Edinburgh handedness scale (Oldfield, 1971). Altogether 10 participants were recorded, 7 females and 3 males, however the data of one speaker had to be eliminated due to some technical problems. Participants received 10 € in compensation for their participation.

The speech material consisted of six common Polish counting out rhymes (Table 1). The participants were instructed to familiarize themselves with the rhymes first and then carry out the tasks as described above.

Table 1: Six Polish counting out rhymes with their orthographic representation, number of syllables (syllables are separated by ".") and words. The ratio is calculated as the No. of syllables/No. of words

Orthographic representation of counting out rhymes	**No. of syll.**	**No. of words**	**Orthographic representation of counting out rhymes**	**No. of syll.**	**No. of words**
Ent.li.czek – pent.li.czek,	6	2	Tre.le.le.le, tre.le.le.le,	8	2
czer.wo.ny sto.li.czek,	6	2	Zja.dłem dzi.siaj trzy mo.re.le.	8	4
na ko.go wy.pa.dnie,	6	3	Raz, dwa, trzy,	3	3
na te.go - bęc!	4	3	Dziś o.bia.du nie jesz ty!	7	5
Ratio = **2**	**20**	**10**	**Ratio = 1.86**	**26**	**14**
Raz, dwa, trzy,	3	3	Bzy, bzy, bzy,	3	3
wy.chodź ty,	3	2	By.ły so.bie pszczół.ki trzy:	7	4
jak nie ty, no to ty.	6	6	Ma.ja, Gu.cio, Kle.men.ty.na	8	3
Ratio=1.1	**12**	**11**	I wy.cho.dzisz ty.	5	3
Pan So.bie.ski miał trzy pie.ski,	8	5	**Ratio = 1.77**	**23**	**13**
czer.wo.ny, zie.lo.ny, nie.bie.ski.	9	3	Wpa.dła bom.ba do piw.ni.cy,	8	4
Raz, dwa, trzy,	3	3	na.pi.sa.ła na ta.bli.cy:	8	3
po te pie.ski i.dziesz ty.	7	5	S. O. S. – głu.pi pies.	6	3
Ratio = 1.69	**27**	**16**	Tam go nie ma, a tu jest.	7	7
			Ratio = 1.71	**29**	**17**

2.3 Data pre-processing, speech and gesture annotations

Motion data were exported to the c3d format. Since we did not use a fixed skeleton, the 15 markers were renamed according to their anatomical position using the Biomechanical Toolkit (Barré & Armand, 2014). Hereafter, a velocity vector (v) of the x, y and z time series with a length from 1 to j was calculated as the central difference for index finger (equation 1).

$$v(j) = \mathrm{sqrt}(((x(j+1)-x(j-1))/2)^2 + ((y(j+1)-y(j-1))/2)^2 + ((z(j+1)-z(j-1))/2)^2) \quad (1)$$

The velocity vector was saved in wav-file format and annotated together with the speech wav-file using Praat (version 6.0.26). In the speech wav-file we manually labelled the on- and offset of the respective counting out rhyme and all silent pauses longer than 100 ms. To calculate speech rate, we divided the number of syllables of the respective rhyme by the summed duration of all speech units without pauses.

Index finger turning points were labelled as velocity minima from the beginning to the end of the counting out rhymes. A stroke was defined as a movement between two successive velocity minima from the speaker pointing towards the teddy bear or back. We removed the first and the last stroke, because – unlike the other strokes – they started or ended in the arm and hand hanging down. For each stroke, the duration, displacement and maximum velocity were calculated.

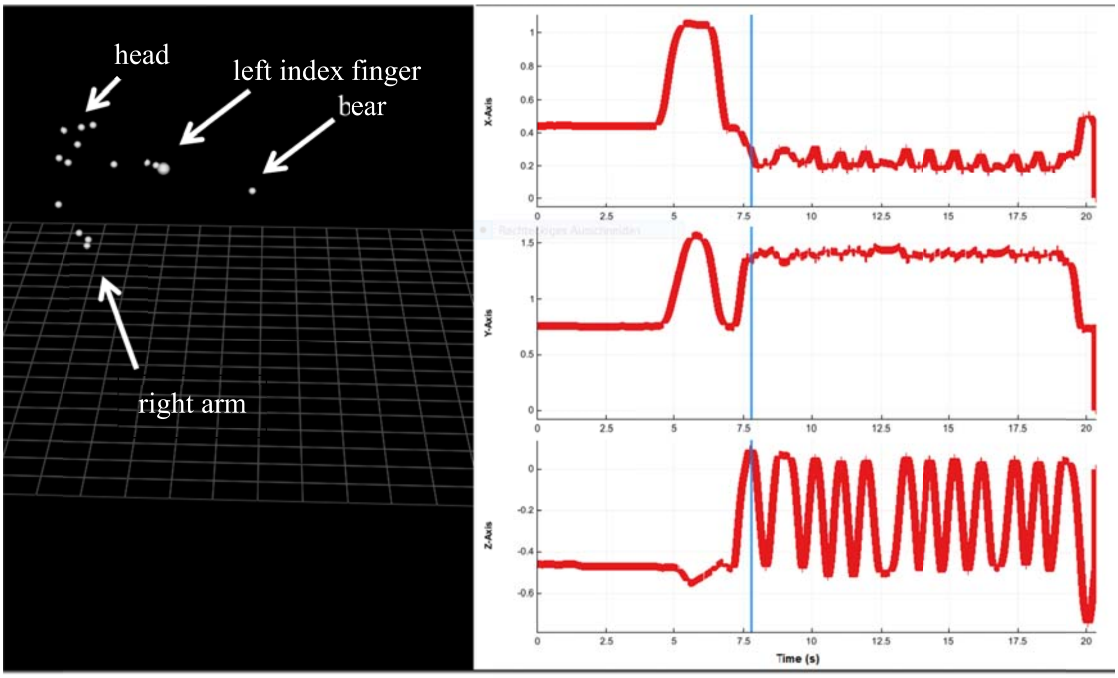

Figure 2. Left: Mokka display of recorded speaker with 15 reflecting markers; Right: x, y and z time series of the dominant index finger.

2.4 Statistical analyses

Since our dataset is balanced, we ran several repeated measures ANOVA to investigate the general effect of CONDITION (normal vs. fast rate) and HANDEDNESS (right hand or left hand) on the ratio of the number of syllables to the number of pointing gestures, the speech rate without pauses, the average stroke duration, average peak velocity and average displacement. Rhymes per speaker were included as an error term. We applied a Bonferroni correction and treated all findings as significant when p was below 0.01 (0.05 / 5, i.e. alpha level of 5% was used and divided by the number of the applied models). All statistical tests were carried out in R (version 3.2.3).

2.5 Results

In order to obtain information about the stability of the relationship between the number of syllables produced and the number of pointing gestures (strokes), we investigated the ratio between the two (Figure 3). Since we did not find any significant differences among the left and right arm, the results are pooled together in the figures. It is evident that speech rate does not lead to a reorganization of the two motor actions and the ratio is rather stable. However, the number of syllables realized within a stroke is highly speaker-specific and ranges from a ratio of 1 (S4, S7) to >2 (S5). Repeated measures ANOVA did not reveal any significant effects for CONDITION, HANDEDNESS or their interaction. Although the ratio is stable, speakers clearly fullfilled the task and increased their speech rate (Figure 4; df=1, F=92.2, $p<0.001$). Moreover, speech rate increasegoes hand in hand with a shorter average duration of the pointing gestures (Figure 5; df=1, F=528.955, $p<0.001$).

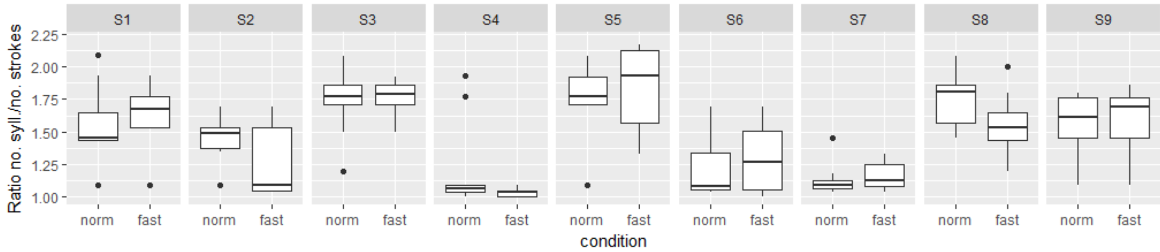

Figure 3. Boxplots showing the relation between number of syllables and number of pointing gestures by speaker in the normal and fast condition.

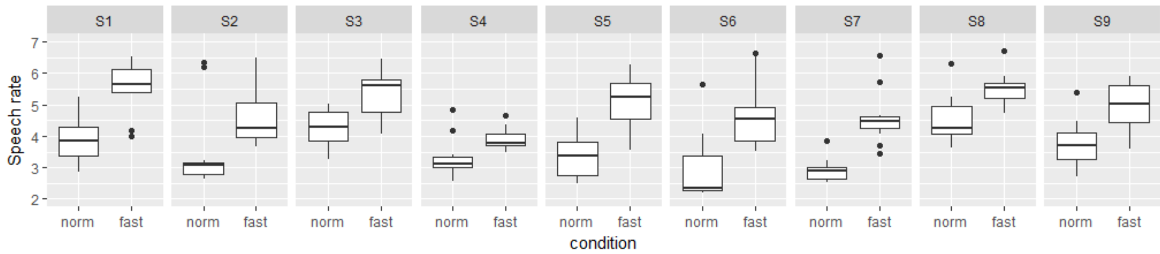

Figure 4. Boxplots showing speech rate (no. of syll/s) split by speaker in the normal and fast condition.

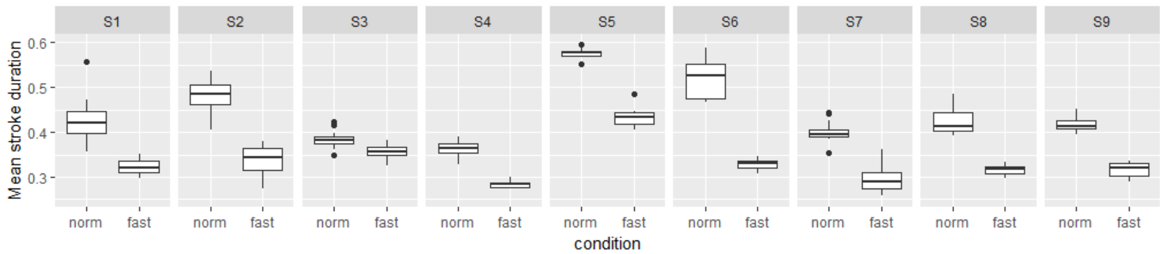

Figure 5. Boxplots showing the averaged stroke duration (s) by speaker in the normal and fast condition.

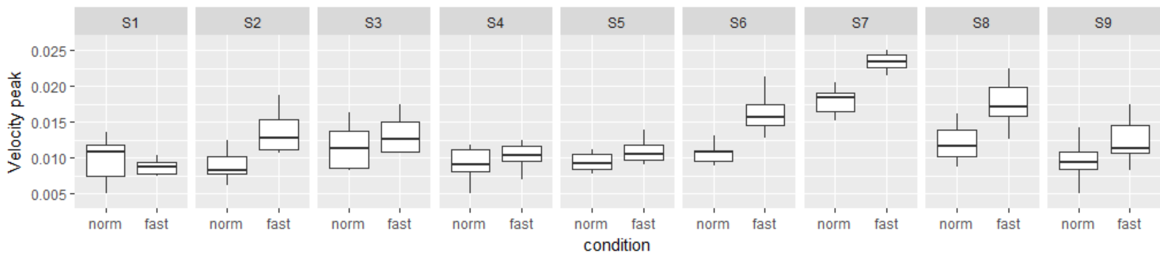

Figure 6. Boxplots showing peak velocity of the stroke split by speaker in the normal and fast condition.

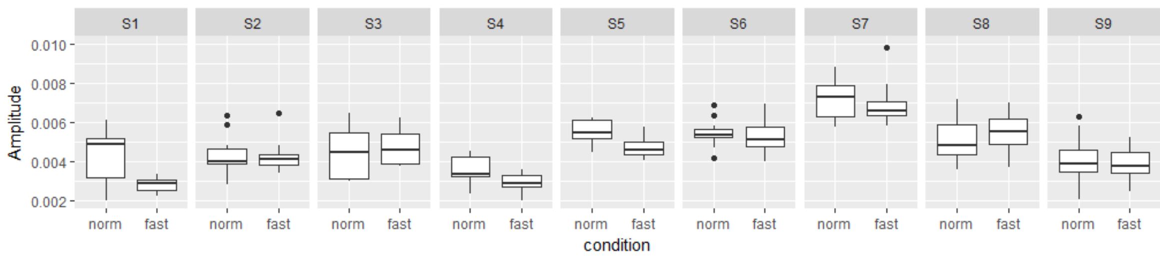

Figure 7. Boxplots showing displacement of the stroke split by speaker in the normal and fast condition.

Finally, pointing gestures are not only shortened, but also produced with a faster velocity of the index finger motion (Figure 6; df=1, F=23.3, p<0.001). Figure 6, however, also displays some speaker-effects. S1, S4 and S5 do not increase the peak velocity substantially, but decrease the distance from one turning point to the next (Figure 7; no sign.).

3 Summary and Discussion

Similar to our previous work on German counting out rhymes (Fuchs & Reichel, 2016), we also found evidence for a relatively stable relation between the number of syllables and pointing gestures for Polish counting out rhymes. The exact relation was, however, speaker-specific. One of the participants of the study even told us spontaneously that he could speak faster, but only without pointing, which further evidences the stable relation and the fact that the slower arm motor system puts some constraints on speech. Hence, physical properties of different motor systems need to be taken into account when studying the relation between speech production and gesturing.

We have also noted some flexibility in rhymes with syllable to word ratios varying considerably from one to the next line. Future work will therefore include a more detailed analyses of each line of the respective rhyme and will carry out a more in-depth coordination analysis.

Stability was occurent independent of whether the dominant or non-dominant hand was used for pointing. In fast speech, most speaker realised similar displacements than in normal speech, but produced a shorter stroke duration and a higher peak velocity. A few speakers kept peak velocity rather stable, but produced smaller movement amplitudes and shorter stroke durations in fast speech. A complete reorganisation was not found, independently of speaker-specificity. In future work, we hope to compare these results with those from different languages.

4 Acknowledgements

This work was supported by a grant from the BMBF (01UG1411) and the Leibniz Society. We would like to thank Olivia Maky for her help in data preprocessing and annotation, Joerg Dreyer for technical support, our participants and Mark Tiede for his scientific input.

References

Barré, A. & Armand, S. (2014) Biomechanical ToolKit: Open-source framework to visualize and process biomechanical data. *Computer Methods and Programs in Biomedicine* 114 (1): 80-87.

Boersma, P. & Weenink, D. (2017). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.26, retrieved 5 February 2017 from http://www.praat.org/

Cummins, F., & Simko, J. (2009). Notes on phase and coordination, and their application to rhythm and timing in speech. *音声研究, Journal of the Phonetic Society of Japan* 13(3), 1-12.

Fuchs, S. & Reichel, U.D. (2016). On the relationship between pointing gestures and speech production in German counting out rhymes: Evidence from motion capture data and speech acoustics. In C. Draxler & F. Kleber (Eds.), *Proceedings of P&P 12*, 1-4. München: LMU.

Hanna, N. A., Patrizia, K. L., & Dufter, A. (2002). The meter of nursery rhymes: universal versus language-specific patterns. *Sounds and systems: studies in structure and change. A Festschrift for Theo Vennemann*, pp. 241-267Berlin/New York: Mouton de Gruyter (Trends in Linguistics).

Iversen, J. R., & Balasubramaniam, R. (2016). Synchronization and temporal processing. *Current Opinion in Behavioral Sciences*, 8, 175-180.

Jannedy, S., Fuchs, S. & Weirich, M. (2010). Articulation beyond the usual: Evaluating the fastest German speaker under laboratory conditions. In S. Fuchs, P. Hoole, C. Mooshammer & M. Zygis (Eds.), *Between the regular and the particular in speech and language*, pp. 205-234. Frankfurt/M.: Peter Lang.

Lancia, L. & Fuchs, S. (2011). The labial coronal effect revisited. In Yves Laprie (ed.), *Proceedings of the ISSP*, 187-194. Montreal: CD-ROM.

Malisz, Z., Żygis, M., & Pompino-Marschall, B. (2013). Rhythmic structure effects on glottalisation: A study of different speech styles in Polish and German. *Laboratory Phonology*, 4(1), 119-158.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsych.*, 9, 97–113.

R Core Team (2015). *R: A Language and Environment for Statistical Computing* [Computer program]. Version 3.2.3., retrieved October 2015 from https://www.R-project.org

Rochet-Capellan, A., Laboissiere, R., Galvan, A. & Schwartz, J.L. (2008). The speech focus position effect on jaw–finger coordination in a pointing task. *JSLHR* 51.6, 1507–1521.