1-7 August 1999. Berkeley: Department of Linguistics, University of California, 1999. p. 399-402.

ARE THERE COMPENSATORY EFFECTS IN NATURAL SPEECH?

Anja Geumann*, Christian Kroos*[†], and Hans G. Tillmann*
*Institut für Phonetik und Sprachliche Kommunikation,
Ludwig-Maximilians-Universität Munich, Germany
[†]ATR Human Information Processing Res. Labs., Kyoto, Japan

ABSTRACT

This work exploited coarticulation and loud speech as natural sources of perturbation in order to determine whether articulatory covariation (motor equivalent behavior) can be observed in speech that is not artificially perturbed. Articulatory analyses of jaw and tongue movement in the production of alveolar consonants by German speakers were performed. The sibilant /s/ shows virtually no articulatory covariation under the influence of natural perturbations, whereas other alveolar consonants show more obvious compensatory behavior. Our conclusion is that an effect of natural sources of perturbation is noticable, but sounds are affected to different degrees.

1. INTRODUCTION

Experiments with artificially perturbed speech (e.g. bite block experiments) show that phonetically defined goals can be reached by different articulatory strategies, i.e. that compensation takes place. The question remains whether this key principle of motor control can also be observed in unconstrained speech. The first problem here arises in identifying natural sources of perturbation. According to Edwards [2] coarticulation can be taken as a natural source of perturbation that influences the interarticulator coordination. The effect of coarticulation on the tongue-jaw interaction in a variety of alveolar consonants was examined in a pilot experiment by Kühnert et al. [4].

A further source of natural perturbation may be found in loud speech [6, 7]. In loud speech the jaw may adopt a more open position, thus forcing a different pattern of interarticulator coordination from that found in speech uttered at a normal volume level.

In our experimental setup we thus decided to use two natural sources of perturbation: coarticulatory effects and loud speech.

2. EXPERIMENT

2.1. Data

Kinematic and acoustic recordings were made of read phrases, produced by 4 German speakers (one female (AW), three male). Pseudo-word 'VCV sequences were embedded in carrier phrases of the type "Hab das Verb ___ mit dem Verb ___ verwechselt".

The target consonants were the alveolar German phonemes differing in manner of articulation /s, ʃ, l, n, d, t/ (/ʃ/ is postalveolar). They were placed in differing symmetric vowel height contexts /i__i, e__e, a__a/; both vowels were long, with main stress on the first vowel. All phrases were produced in loud and normal speech, which was elicited by simple instruction of the speaker. The loud and normal phrases were presented in

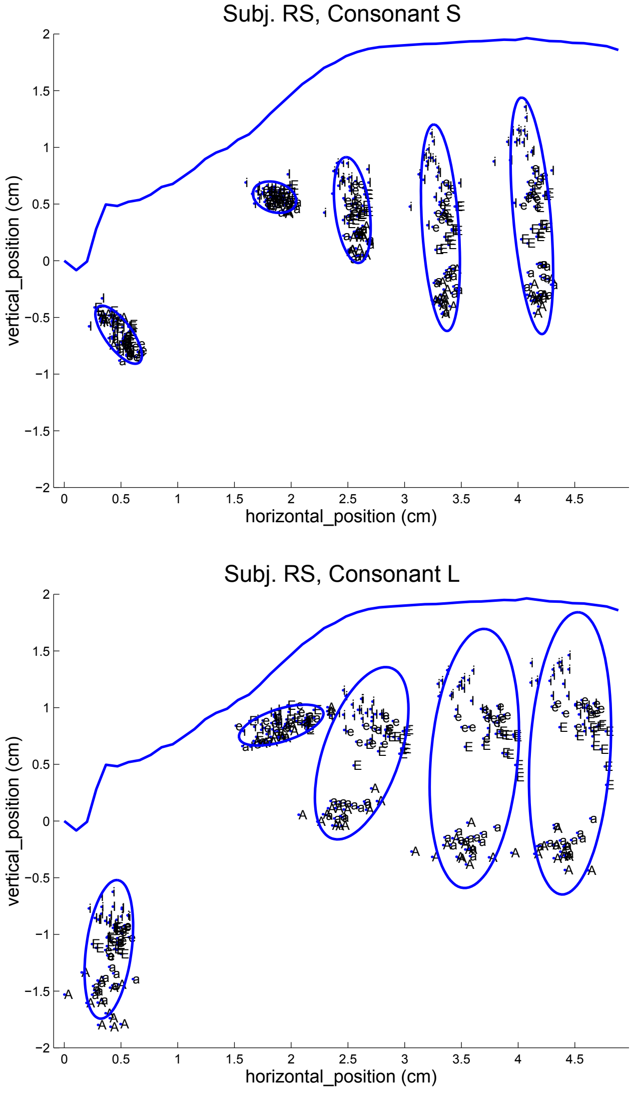

Figure 1. Articulatory data for speaker RS for /s/ and /l/. The symbol indicates the vowel context, with i, e, a = normal volume; I, E, A = loud volume. Anterior is to the left. Sensors from left to right: *jaw-out*, *tongue-tip*, *blade*, *dorsum*, *back*. Radius of main axis of ellipse equals twice the standard deviation along the first principal component of variation.

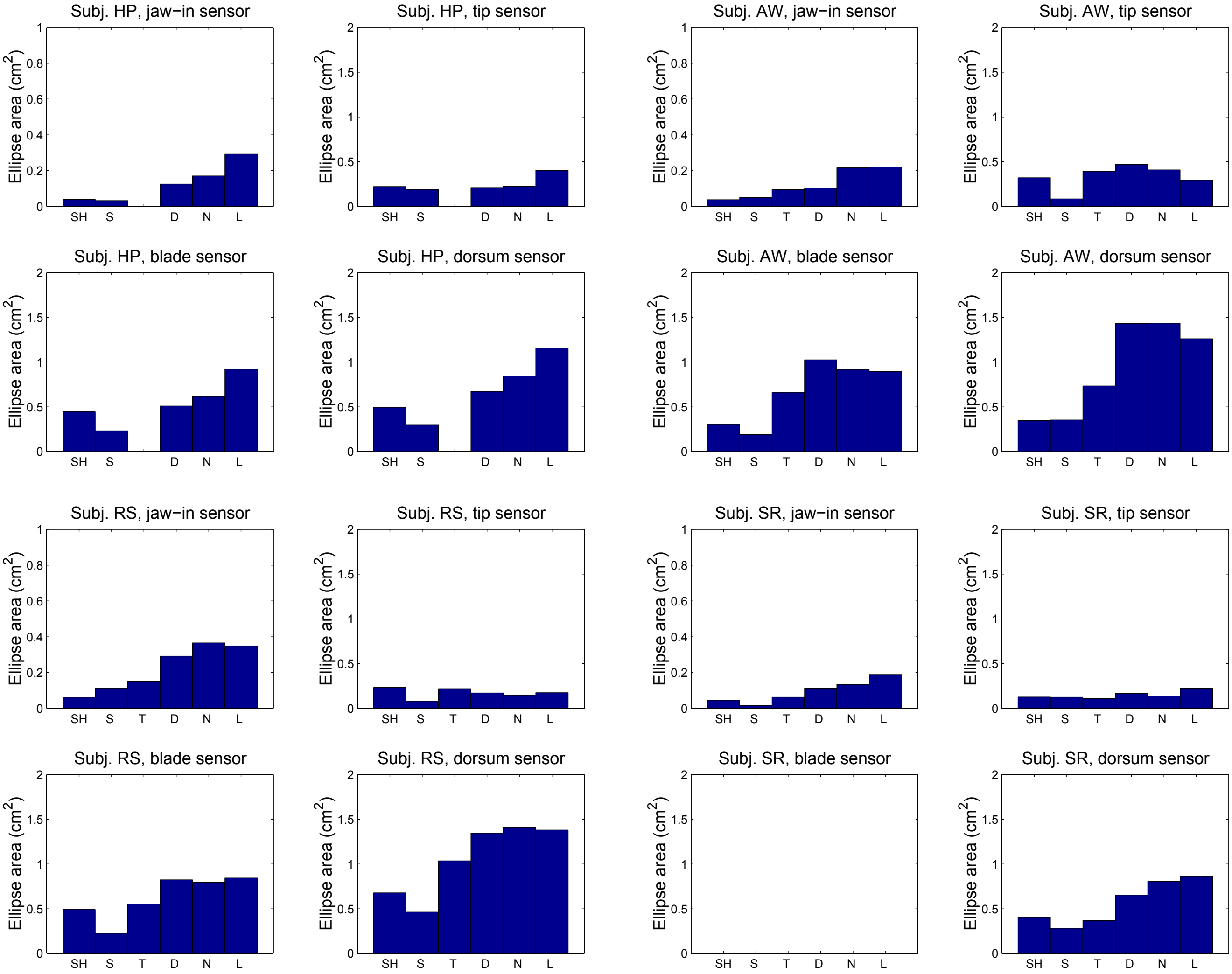

Figure 2. Articulatory variability of target consonants over all speech volumes and vowel contexts. Separate panels for each subject (HP, AW, RS, SR). and one jaw and three different tongue sensors (*jaw-in*, *tip*, *blade*, *dorsum*) The *tongue-back* sensor has been omitted but shows results similiar to *dorsum*. For speaker HP a corpus was used that did not contain the consonant /t/; for speaker SR the tongue-blade sensor failed during the experiment.

random order. For each target consonant with given loudness and vowel context 12 repetitions were produced, i.e. 72 repetitions of each consonant over all context and loudness conditions.

The two-dimensional (midsagittal) kinematic signals were recorded with an electromagnetic transduction system (Articulograph AG100, Carstens Medizinelektronik, for more technical details see Hoole [3]). Four sensors were placed on the tongue (referred to as *tip*, *blade*, *dorsum* and *back)*: The tip sensor was placed approx. 1cm posterior to the tongue tip, and was assumed to best track alveolar articulation. The other three followed in equidistant steps up to a point opposite the junction of hard and soft palate (*blade*, *dorsum*, *back*). Three sensors were used to track the jaw movement. One each was placed on the inner (*jaw-in*) and outer (*jaw-out*) surface of the gums beneath the lower incisors, a third sensor was placed on the angle of the chin (*chin*). Reference sensors were located on upper jaw and the nasion.

2.2. Results

Three analysis points during the acoustic manifestation of the target consonant were determined: First, the acoustic midpoint of the consonant, second, the point of minimal tangential velocity of the tongue-tip sensor trajectory; and third the point of minimal tangential velocity of the *jaw-out* sensor trajectory. Comparison of these three points showed little effect on the results. Thus, in the following, the analysis for the acoustic midpoint alone is presented.

In figure 1 the overall variability in the articulatory data for one speaker is illustrated. In our experiment loudness and coarticulation are essentially intended to evoke gradually varying effects on jaw height. Differences between the different sources of perturbation will not be considered further here.

2.2.1. Articulatory variability for different sounds and sensor positions. The data in figure 1 show a relatively continuous pattern of variation in jaw height for /l/ whereas for /s/ only a very reduced variability is found. In figure 2 a more schematic overview of the variability for all sounds and all speakers is presented. Three outcomes of the data presented in figure 2 can be summarized: 1. The variability in the tongue position increases with distance from the alveolar place of articulation. 2. The fricative /s/ exhibits least variability over all sensor positions. 3. The differences in variability between /s/ and the more variable consonants like /l/ and /n/ are stronger at the back parts of the tongue (i.e. at those remote from the place of articulation).

2.2.2. Complementary covariation between tongue tip and jaw. For the next step we examined the interaction between jaw height and tongue tip height. As is well known the two articulators are not independent of one another. The measured tongue tip position consists of a real (henceforth intrinsic) tongue tip position and a share of jaw movement.

The intrinsic tongue height here was estimated by simple subtraction of the y-value of *jaw-in* from the y-value of *tongue-tip*. (cf. [4]). The results of *jaw-in* were similiar to those of the other jaw sensors but provided more conservative results. Figure 3 shows detailed results for two contrasting speakers, RS and AW. The negative correlations for speaker AW are much weaker than those of RS. In AW as well as in RS the strongest negative correlation can nonetheless be found in /l/ and /n/, the weakest in /s/.

In figure 4 a different way of describing the effect of covariation is presented. It summarizes the correlation coefficients for all speakers and consonants and relates these results to the standard deviation of the jaw. On the one hand this figure demonstrates more clearly that for /s/ production the standard deviation of the jaw height is only about 0.5mm or even less for three of the four speakers. For the remaining speaker RS the jaw height variation of /s/ is also relatively small compared to the other sounds. Congruously only weak covariation for the /s/ can be expected. More reliable candidates for tip-jaw covariation are those that show a larger variation in jaw height combined with a high negative correlation between jaw height and intrinsic tip height. For all four speakers /n/ and /l/ show up with this tendency.

The problem of factoring out intrinsic tongue height by subtraction of the jaw and afterward correlating these (dependent) data with jaw height (cf. [1]) should not be underestimated.

One major advance of our experiment in contrast to the experiment by Kühnert et al. [4] is that we are already able to compare three different sensors that monitor jaw movement. The risk of overestimating the negative correlation between the coupled articulators can be decreased by choosing the sensor giving the weakest negative correlations.

Full estimation of intrinsic tongue activity requires factoring out the jaw contribution to measured tongue position. This requires, in turn, decomposition of jaw movement into rotational and translational components. We are at present working out an operation to do this decomposition by combining the information of the three jaw sensors with anatomical information (retrieved by means of NMRI) about position of the condyle in our subjects.

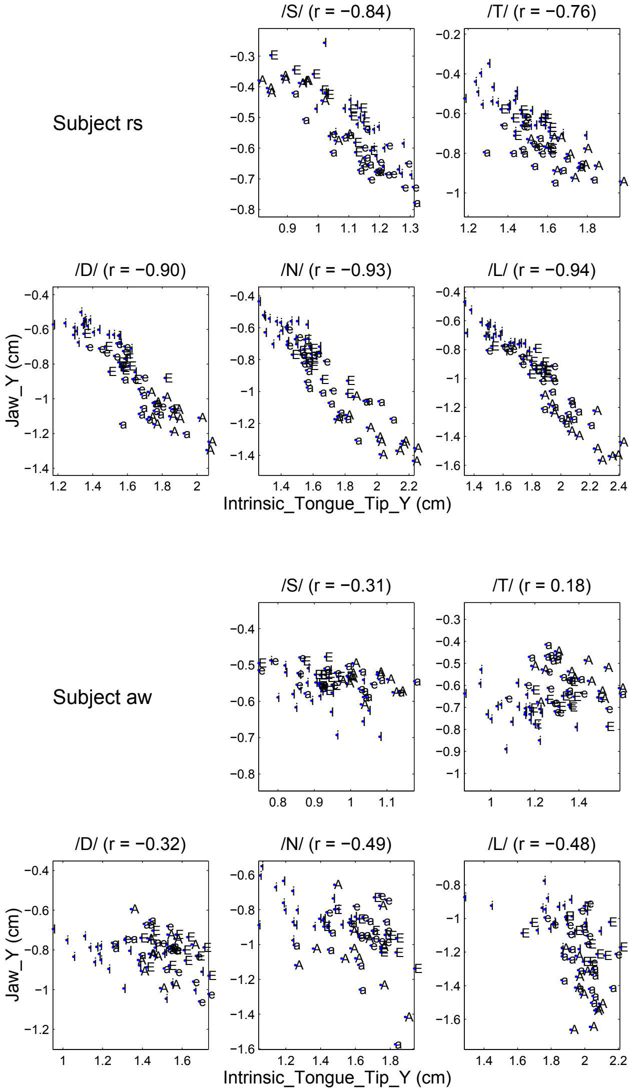

Figure 3. Covariation between jaw height and intrinsic tongue height for speakers RS (top) and AW (bottom). Symbols as in figure 1. Results for /ʃ/ have been omitted.

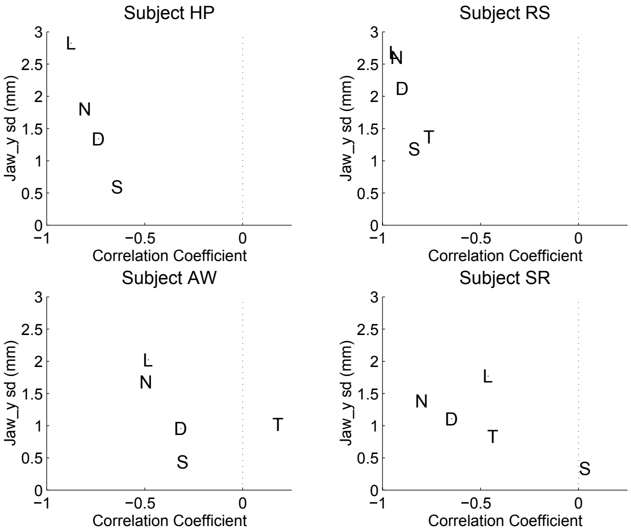

Figure 4. Abscissa: Correlation coefficient between jaw height and intrinsic tongue height; Ordinate: Standard deviation of jaw height. Subjects HP, RS, AW, SR.

3. CONCLUSION

Our expectation was that V-to-C coarticulation and different loudnesses would cause the jaw position in consonants to vary. We supposed further that speakers would use the tongue for compensation to keep vocal tract constriction in the consonants relatively constant. Specifically we were interested in whether trade-off effects (strength of complementary covariation) would vary over consonants sharing place of articulation but differing in manner. One preliminary hypothesis, motivated by one speaker in the experiment of Kühnert et al. [4], was that trade-offs would be most apparent in acoustically sensitive sounds such as fricatives. This was not confirmed by our experiment. Jaw position was so precise for the fricatives that no lingual compensation was required. Nevertheless, for both jaw and tongue, variability increased from fricatives via stops to the lateral and nasal. All the same it may be oversimplifying to merely state that /l/ is more variable than /s/. Even in /l/ (compare figure 1) the variability in the tongue tip is relatively small compared to the variability of the back parts of the tongue. This leads us to assume that especially for /s/ the jaw, as well as the tongue tip, plays the role of an articulator whose precise positioning is crucial. The important role of precise positioning of the incisors and thereby the jaw for the production of sibilants has been pointed out by Shadle [8, 9] and subsequently Ladefoged [5].

This account explains the patterns we found in our speakers and the patterns observed in a similiar experiment [4] for two of three speakers.

This leaves one speaker in [4] who was reported to show a pattern more or less opposite to our present findings. So probably - after all - individual strategies even for /s/ may vary.

The question raised in the title of this work remains. Our findings indicate that the paradigm of motor equivalence as a genuine feature of natural speech requires a sound-specific perspective. Before judgement is possible of what can be compensated there has to be a more complete account of sounds in terms of relevant articulatory goals and more irrelevant properties. Further, although we have tried to retrieve our results out of a corpus of relatively natural speech with natural sources of perturbation, a necessary further step will be for us to obtain data from a more natural corpus with the target sounds embedded in real German words. This work is currently underway.

ACKNOWLEDGMENTS

This work was funded by the German Research Council (DFG) with a grant to the third author (Ti69/31). This paper would have had a fourth author: Phil Hoole; but unfortunately the number of his presentations at this conference is limited to two.

REFERENCES

[1] Benoit, C. 1986. Note on the use of correlations in speech timing. *Journal of the Acoustical Society of America* (*JASA) 80*, 1846-1849.

[2] Edwards, J. 1985. Contextual effects on lingual-mandibular coordination. *Journal of the Acoustical Society of America* (*JASA) 78*, 1944-1948.

[3] Hoole, P. 1996. Issues in the acquisition, processing, reduction and parameterization of articulographic data, *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation, München (FIPKM)*, 34, 158-173.

[4] Kühnert, B., Ledl, C., Hoole, P. & Tillmann, H.G. 1991. Tongue-jaw interactions in lingual consonants. *PERILUS (Phonetic Experimental Research at the Institute of Linguistics University of Stockholm) 14*, 21-25.

[5] Ladefoged, P. 1990. On dividing phonetics and phonology: comments on the papers by Clements and by Browman and Goldstein. In: J. Kingston & M. Beckman (Eds.) *Papers in Laboratory Phonology 1*, Cambridge: Cambridge University Press, 398-405.

[6] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H & H theory. In: W. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 403-439.

[7] Schulman, R. 1989. Articulatory dynamics of loud and normal speech. *JASA, 85*, 295-312.

[8] Shadle, C.H. 1985. The acoustics of fricative consonants. Ph.D. thesis, MIT.

[9] Shadle, C.H. 1990. Articulatory-Acoustic Relationships in Fricatice Consonants. In: W. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 187-209.