Gick, B., Iskarous, K., Whalen, D. H., & Goldstein, L. M. (2003). Constraints on variation in the production of English /r/. In S. Palethorpe & M. Tabain (Eds.), *Proceedings of the 6th International Seminar on Speech Production* (pp. 73-78).

1327

CONSTRAINTS ON VARIATION IN THE PRODUCTION OF ENGLISH /r/

Bryan Gick [1, 2], Khalil Iskarous [2],

D. H. Whalen [2] & Louis Goldstein [2, 3]

[1] Dept. of Linguistics, University of British Columbia, Canada
[2] Haskins Laboratories, New Haven, USA
[3] Yale University, USA

ABSTRACT : English /r/ has been shown to vary across dialects and speakers, and even within speaker. One account for this variation argues that acoustic values are stable despite substantial articulatory variations, resulting in a "trading" relationship favoring acoustics over articulation (Guenther & al. 1999). This paper questions the evidence and need for this acoustic representation, arguing that acoustic variation follows articulation. First, correlations in kinematic point-tracking patterns previously raised in defence of acoustic-articulatory trading relations are found to occur not just in /r/, but across other segments lacking /r/'s proposed F3 acoustic "target", suggesting more general physical constraints on tongue kinematics. Second, using a combination of ultrasound and video imaging, all three component gestures (lips, tongue anterior, tongue root) for English /r/ are observed to be consistent with a model of articulatory targets specifying only constriction locations, suggesting that any stability in the acoustic signal corresponds with these stable articulatory parameters. It is further suggested that articulatory differences previously observed in /r/ variants are overemphasized by point-tracking measurement devices, which are unable to measure actual constriction locations as directly as imaging techniques.

INTRODUCTION

English /r/ has long been known to show a high degree of variability in production across dialects and speakers (e.g., Delattre & Freeman 1968, Westbury et al. 1998), and even across context within speaker (e.g., Guenther et al. 1999). One attempt to account for this variation argues that acoustic values - in this case, an F3 target - may be stably maintained despite substantial articulatory variations, resulting in a "trading" relationship favoring acoustics over articulation (Guenther et al. 1999). The present paper reexamines these /r/ variants, questioning whether there exists sufficient evidence to suggest that acoustic targets are indeed more stable than articulatory ones.

A number of papers have sought evidence for acoustic invariance via acoustic-articulatory tradeoffs (e.g., Perkell et al. 1993, Guenther et al. 1999). This research program has been based on the premise that, if one kind of speech event is found to show invariance in one domain (say, acoustics), but not the other (say, articulation), then all speech events should be described using this (more invariant) domain. If this is true, then one need only look as far as the many examples of articulations that consistently occur without any acoustic realization at all (e.g., the well-known case of "perfec[t] memory" [Browman & Goldstein 1990, replicated under varying rate conditions by Tiede et al. 2001]) to demonstrate that acoustic targets cannot possibly be the goals of speech production. Whether or not this is a sound premise, however, it is not the purpose of this paper to demonstrate that articulatory goals are invariant and are thus the "true" targets of speech production. Rather, it is our goal to show that: 1) articulatory variance is adequately described in terms of physical factors, and 2) articulatory variance, where it occurs, is represented in acoustics.

PHYSICAL FACTORS INFLUENCING KINEMATIC PATTERNS

Guenther et al.'s (1999) analysis of /r/ rested upon linking correlations found in the movement of tongue points to the maintenance of a relatively stable F3 target. The correlations they observed were:

1) a positive correlation between tongue back height and tongue front horizontal position (horizontal values increase with backing and decrease with fronting): found for 7/7 subjects

2) a negative correlation between tongue back height and tongue front height: found for 5/7 subjects

3) a positive correlation between tongue front horizontal position and tongue front height: found for 1/7 subjects

These correlations are interpreted as articulatory trading relations that maintain a stable F3. However the existence of a consistent relation between the motion of different points of the tongue could be the result of the fact that the points are physically linked. The tongue is a hydrostatic structure, therefore the motion of its points are expected to be highly correlated. Previous studies of acoustic-articulatory trading relations in articulation have not addressed the possibility that the trading relations may be the result of general constraints on tongue movement like those imposed by its hydrostatic structure. One way to test for overriding physical constraints on the tongue is to test for similar movement correlations in sounds not expected to maintain a stable F3 target. Specifically, we suggest that the first two correlations are consistent with a volume constraint on the hydrostatic tongue, and that the third constraint is consistent with limitations of the palate angle on tongue position. These are illustrated in Figure 1. Thus, for example, simultaneous raising of the tongue dorsum and fronting of the tongue tip would imply a general expansion of the volume of the tongue, and should therefore be expected not to occur, predicting the correlation shown in (1).

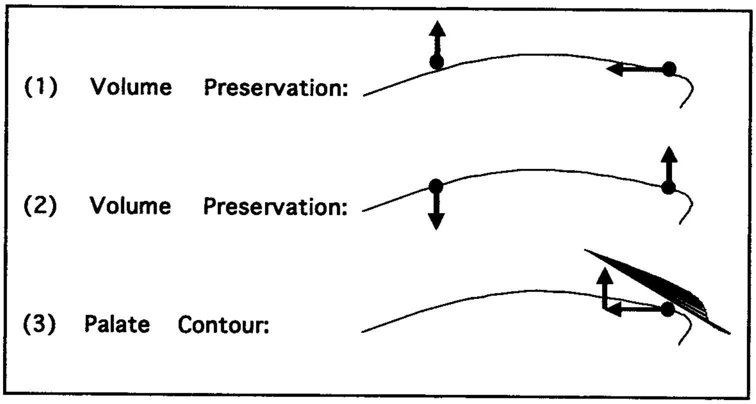

Figure 1. Illustration of proposed effects of physical constraints on tongue movement, corresponding with correlations from Guenther et al. (1999)

An experiment was conducted to test whether these same correlations occur in high front vowels, which most closely resemble /r/ in the anterior region under investigation (for testing of continuous speech and other vowels, see Brown & Gick 2003).

Methods

Data for this experiment were extracted from the University of Wisconsin X-ray Microbeam Database (Westbury 1994). Kinematic records for 3 arbitrarily chosen subjects (the first 3 in the database) were analyzed. Positions were recorded from the midpoints of all available non-low front vowels (26, 29 and 25 vowels for subjects 1, 2 and 3, respectively) for transducers attached to the tongue anterior and posterior points (T1 and T3, numbered front-to-back) corresponding most closely with those used by Guenther et al. (1999). Correlations between transducer positions were tested for significance using a Pearson's correlation matrix.

Results

Table 1 reports the statistical results, and Figure 2 shows a scatterplot of the results for each subject for each correlation tested. Results show a significant positive correlation for all 3 subjects for the Correlation 1 (T1x vs. T3y), a significant positive result only for S2 for Correlation 2 (T1y vs. T3y), and no significant results for Correlation 3.

Table 1. Results and significance levels of Pierson's correlations for all subjects and all correlations.

Correlation	Cor 1			Cor 2			Cor 3		
Subject	S1	S2	S3	S1	S2	S3	S1	S2	S3
n	26	29	25	26	29	25	26	29	25
r	.570	.425	.504	.270	.784	.297	-.266	.285	-.336
p	<.01	<.05	<.01	NS	<.01	NS	NS	NS	NS

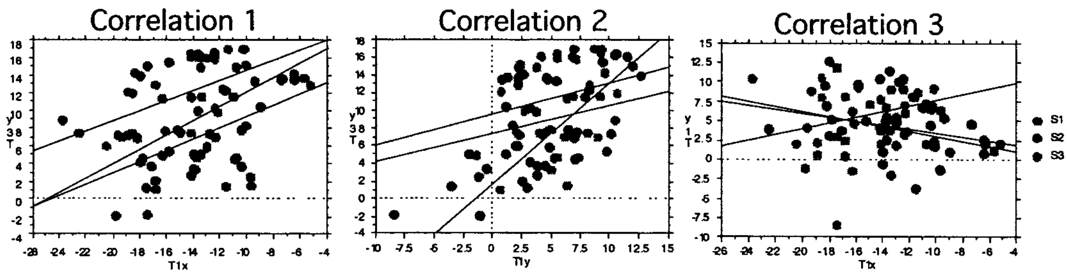

Figure 2. Scatterplots for each of the correlations tested on high front vowels. Regression lines separated by subject.

Discussion

The above results indicate that high front vowels share with /r/ the significant positive correlations for tongue back height and tongue front horizontal position. Here, as in Guenther et al.'s (1999) study, this was the only correlation found to be significant across all subjects. The second correlation was found to be significant only for one subject, and in the opposite direction from that predicted, while the third correlation was not significant for any subject. These findings suggest that the main correlation observed for /r/ by Guenther et al. (1999) is the result of a physical constraint on the tongue – possibly a volume constraint – and not of an acoustic F3 target. Further testing of additional subjects is under way.

VARIANCE AND INVARIANCE IN /r/

The story about /r/ that has now been told many times is that there is a many to one relation between /r/ articulation and acoustics, and therefore the acoustic invariant of low F3 ought to be taken as the target for that segment. We take issue with this account. There is of course no question that articulatory variability, largely determined by coarticulatory and planning factors, exists in the production of /r/, but we argue that this articulatory variability results in acoustic variability as well, so that the variability exists in both domains. Moreover, the commonly observed acoustic feature of low F3 is matched on the articulatory side by the three gestures of /r/ that are present in the different reflexes. Variability and invariance therefore exist in both domains.

The notion that the different variants of /r/ are acoustically equivalent results from a focus on static measurements of F3. Recent results suggest that the /r/ variants are not acoustically equivalent. In an EMMA study using 5 subjects, Espy-Wilson and Boyce (1999) found that static measurements of F4 correlate with the different articulatory variants of /r/. The static articulatory variation is therefore reflected at the acoustic level. Also, in a study using data from the X-ray Microbeam Database, Hashi et al. (2003) have found a strong correlation between the motion of points of the tongue during the different variants of /r/ and the slope of F3. Further, figure 3 shows the results for one subject in a pilot study of /r/ production (Gick 2002) wherein tip-up and tip-down variants are forced by optimization in the planning of complex sequences of alveolar flaps. Tongue tip position was confirmed using ultrasound imaging. For this subject, static F1, F2 and F3 were all significantly different for the two variants, with F3 showing the greatest magnitude effect. This and the previous results show that articulatory variation is clearly carried over into acoustic variation. The argument could be made in response to these findings that the acoustic differences between the /r/-variants do not matter, since they do all share a relatively low static F3, which can be seen as the /r/-cue, rendering the variants equivalent. However by the same logic, /i/ and /u/ might be considered the same vowel, as they share a relatively low F1.

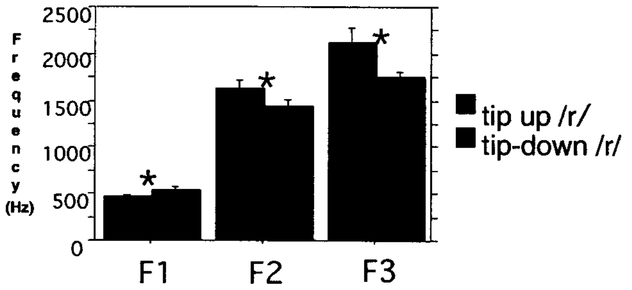

Figure 3. Mean static acoustic centerpoint measures for tip-up and tip-down /r/ variants. Asterisks indicate significant differences (adapted from Gick 2002).

The variants of /r/ are articulatorily and acoustically different. Why then call all of them /r/? The answer has been known as far back as Delattre & Freeman (1968). The variants of /r/ share a labial, palatal, and pharyngeal gesture. The main articulatory difference is in the part of the tongue used to make the palatal constriction, which can vary from the tongue tip to the palatal portion of the tongue, depending on coarticulatory, planning, and perhaps other factors. That is, what differentiates the variants of /r/ is which part of the front of the tongue is used to execute the palatal constriction, but what is articulatorily invariant amongst them is that they all have a constriction in that region, as well as the pharyngeal and labial constrictions. Indeed the invariant palatal gesture can be seen in the data of Guenther et al. (1999). In Figure 17 of that paper, the palatal constriction is shown for all seven speakers in the study. Despite the fact that some subjects use the retroflex variety and others use the bunched variety (as well as variants in between), all the subjects achieve a palatal constriction. The similarity between these two /r/ variants, we argue, is obscured by the use of point-tracking measurement techniques, which focus on articulator positions rather than constrictions. This problem can be avoided by using imaging rather than tracking techniques, particularly where acoustic-articulatory relations are under investigation.

Thus, on the articulatory side, the variants of /r/ are united by having the same constrictions, and are differentiated by the region of the tongue used to make the tongue constriction, while on the acoustic side, the gestures of /r/ result in what is acoustically relatively invariant about /r/, namely the lowered F3 due to the palatal and labial constrictions. The fact that different tongue portions are used in the variants is acoustically reflected in several studies in the form of a different time course for F3 as well as in the static differences observed in F1-F4.

CONCLUSION

Findings of this paper suggest that:

1) Articulatory maneuvers of the tongue for r are not dictated by the achievement of an acoustic invariant, but by physical constraints on the tongue and the articulatory task of the tongue in the production of whichever variant of r is being produced.

2) There is no articulatory-acoustic mismatch for r: The variables and invariants of production are acoustically interpreted.

The account we have provided for American English /r/ may also make sense of what is known about the difficulty in the learning of /r/ by children and SL learners of English (Shriberg 1993, Yamada and Tohkura, 1992). There are at least 3 gestures that could lower F3, labial, palatal, and pharyngeal. These gestures could be combined in many different ways to achieve F3 lowering. If the goal in /r/ production were to lower F3, the speaker has before her a wide variety of choices, and she could choose the one that makes her goal easy to achieve in the particular context. The prediction, therefore, is that /r/ is an "easy" segment, since its goal can be achieved in so many ways. But /r/ is certainly not "easy". If the goal in /r/ is the concurrent production of 3 gestures, we have an immediate account of the difficulty. Few segments have as many simultaneous supra-laryngeal gestures.

To place these findings in the context of other research, studies involving perturbation of auditory feedback (Houde & Jordan 1998, Jones & Munhall 2000) show that at least some aspects of production are modulated by auditory feedback over long durations of continuous speech. However, as Jones & Munhall also show, the effects of this feedback having altered our articulatory targets linger after removal of the perturbation. Thus, consistent with the observations of Browman & Goldstein (1990) and Tiede et al. (2001), we maintain a model whereby the immediate targets of speech production are articulatory, but where we are able to access our metalinguistic knowledge of our own production, allowing auditory feedback to be used to adjust these targets over time, enabling us to respond to changing conditions in our physical states and surroundings. In this way, articulatory and acoustic variability should occur in tandem in unperturbed speech, and both are necessary components in the execution of speech.

ACKNOWLEDGEMENTS

The authors wish to thank Jason Brown, Fiona Campbell, Frank Guenther, Shaffiq Rahemtulla, Ian Wilson, and audiences at the Cancun and Nashville ASA meetings for helpful contributions and discussion, and to acknowledge the generous support of an NSERC Discovery Grant to the first author, and NIH Grant DC-02717 to Haskins Laboratories.

REFERENCES

Espy-Wilson, C. Y., & Boyce, S. E. (1999) "A simple tube model for American English /r/" *Proceedings of the XIVth International Congress of Phonetic Sciences.* J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.). Berkeley, CA: University of California, Berkeley. 2137–2140.

Browman, C. P. and Goldstein, L. (1990) "Tiers in articulatory phonology, with some implications for casual speech" *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* J. Kingston and M. E. Beckman (Eds.). Cambridge: Cambridge University Press. 341-376.

Brown, J. and Gick, B. (2003) "The effect of physical constraints on articulatory variability in English /r/" *Journal of the Acoustical Society of America* **113(5)Part2**: 2258.

Delattre, P. C. and Freeman, D. C. (1968) "A dialect study of American *r*'s by x-ray motion picture" *Linguistics* **44** 29-68.

Gick, B. (2002) "Exercises in 'absurdity': Interactions between flapping and tongue position in American English with implications for speech motor planning" *Journal of the Acoustical Society of America* **112(5)Part2**: 2416.

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M. and Perkell, J. S. (1999) "Articulatory tradeoffs reduce acoustic variability during American English /r/ production" *Journal of the Acoustical Society of America*, **105(5)**: 2854-2865.

Hashi, M., Honda, K., and Westbury, J. (2003) "Time-varying acoustic and articulatory characteristics of American English [r]: a cross-speaker study" *Journal of Phonetics* **31** 3-22.

Houde, J. F. and Jordan, M. I. (1998) "Sensorimotor adaptation in speech production" *Science* **279(5354)** 1213-1216.

Jones, J. A. and Munhall, K. G. (2000) "Perceptual calibration of F0 production: Evidence from feedback perturbation" *Journal of the Acoustical Society of America* **108** 1246-1251.

Perkell, J. S., Matthies, M. L., Svirsky, M. A. and Jordan, M. I. (1993) "Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot 'motor equivalence' study" *Journal of the Acoustical Society of America* **93**: 2948-2961.

Tiede, M. K., Perkell, J., Zandipour, M., and Matthies, M. (2001) "Gestural timing effects in the 'perfect memory' sequence observed under three rates by electromagnetometry" *Journal of the Acoustical Society of America* **110**: 2657.

Westbury, J. R., Hashi, M. and Lindstrom, M. J. (1998) "Differences among speakers in lingual articulation for American English /r/" *Speech Communication* **26**, 203-226.