Articulator movements in ventriloquists' speech

John R. Westbury and **Clarissa J. Weiss**
Department of Communicative Disorders, University of Wisconsin, Madison, Wisconsin, USA
westbury@wisc.edu, cjweiss3@wisc.edu

ABSTRACT

Tongue, lip and jaw flesh-point movements were recorded together with the speech wave from two American English-speaking ventriloquists, each talking normally and then practicing their craft. These recordings were analyzed to describe and quantify changes in articulation arising from self-imposed perturbations of lip and jaw movements during ventriloquism. Results show a variety of apparent compensations involving changes in magnitude and timing of movements, and the positions of articulators. Some compensations are easy to understand from the perspective of a theory of speech production in which acoustic properties of speech sounds are a primary goal. Information from this study may have a practical benefit for disordered patients (e.g., with localized facial paralysis) who must maintain speech intelligibility in the context of diminished capacity for articulatory movements.

1 INTRODUCTION

Ventriloquists speak without appearing to do so, often through an animated puppet or character whose voice quality may be different from their own. This speech illusion requires strong constraints on usual movements of the face and jaw, and hence compensatory articulations for sounds for which these structures are typically involved, if the talker is to sound relatively natural and be well understood. One theoretical reason to study and compare ventriloquial or "vented" speech with normal articulation is to gain insight about control principles and goals underlying production. The basis for the comparison is straightforward: If we discover what talkers preserve in speech after normal articulation has been constrained, we can then infer what they might be trying to do when they speak normally, and possibly, how and why they control their articulators as they do. Certain well-known experimental attempts to perturb normal articulation using bite blocks [3] and lip tubes [9] represent this line of thinking. Results from such studies are commonly cited to show that talkers preserve the sound of speech when they can, creating area functions that approximate their normal ones in spite of, and often in rapid response to novel perturbations of normal articulation. Vented speech involves articulatory constraints that are clearly different from such experimental perturbations. Venting constraints are neither novel for the talker nor especially transient, and the talker's responses to them are highly practiced. Nonetheless, there is a common spirit linking the study of ventriloquism with perturbation experiments, and even with comparative analyses of normal token-to-token variation in articulation and speech acoustics [4, 8]. In broad terms, such studies give us insight about what talkers do with their articulators, and apparently what they want to do, to be understood as they speak.

A second and possibly more practical reason to study and compare ventriloquism with normal speech is to learn about strategies talkers might apply to compensate for functional disorders that selectively affect some articulators. As an example, we know that venting talkers voluntarily solve a problem something like that faced by patients with localized facial paralysis (e.g., associated with bilateral Bell's palsy, or the rarer Möbius syndrome[12]). Any practiced ventriloquist can produce good-quality, intelligible speech despite highly constrained lip movements. It is possible that knowing how they do so might benefit patients affected by these neuromotor deficits. Partly for this practical reason, and for the theoretical reason noted previously, we set out to learn whether and how talkers move their articulators differently during vented than normal speech.

2 METHODS

The X-ray microbeam technique was used to track movements of small markers attached to the tongue, lips, and lower jaw during normal and vented speech. Gold pellets (~ 2.5-mm in diameter) were glued in the midline (a) along the central sulcus of the tongue (~ 1, 2, 4, and 5.5 cm back from the apex), and (b) at the vermillion border of each lip; and, at the incisors and a left-side molar tooth of the jaw. Sagittal-plane positions of these eight markers were sampled in real time, 40-160 times/second, as our talkers moved to speak. Marker movements were recorded synchronously with the radiated speech sound pressure wave sampled ~22,000 times/second. A variety of post-processing and representational conventions applied to these data can be found elsewhere [11].

The talkers in our study were both native, monolingual speakers of American English. One was a 51-year-old retired police officer from the US upper Midwest who used typical dyadic ventriloquism (performer + character) in educational programs for children. The other talker was a 24-year-old female student from the southwestern US who used three-part ventriloquism (performer + character + character's character) in talent exhibitions. The male talker had relatively many dental fillings, and these often caused mis-tracking especially of the three more dorsal tongue markers (T2 - T4). The female talker had fewer dental fillings, and provided correspondingly more complete, better quality data.

Speech materials recorded from both talkers were organized into short records or recording windows, 9 -27 seconds long. During records, talkers read either from (a) lists of isolated words or nonsense (di)syllables, each repeated citation-style, separated from adjacent items in a list by a short pause; (b) lists of short sentences, read at a

ISBN 1-876346-48-5

comfortable, self-selected rate, again with each sentence separated from others in the list by a short pause; or, (c) connected-speech passages ~ 20 seconds in length, each representing part of the *Hunter Script* [1].

Speech tasks produced with and without markers in place were recorded from both talkers speaking as themselves; from the male venting in his character's voice; and, from the female venting in her own voice. A set of three sentence-type records was also recorded from the female with markers in place, venting as her main character. For each markers-on speaking condition (i.e., normal or vented), there were 13 word-type records, containing 56 different real words and 118 total word-like tasks; 14 sentence-type records, containing 10 different sentences and 34 total sentences; and, 4 of the connected speech type. Added together, these records yielded ~14 minutes of speech tracking data from each talker. A subset of the markers-on task set – including 9 word-type records, and 5 sentence-type records – was recorded from both talkers speaking normally and venting but without markers in place. Thus, both talkers served as their own experimental controls, producing both normally articulated and vented tasks, with and without markers.

A subset of tasks from the inventory was selected as foci in our initial comparative analysis of normal and vented speech. These included (a) the list-final foil word *seed*, because it was repeated more often than any other word, and because we expected its articulation would **not** be much affected by venting; (b) the four connected speech passages forming the *Hunter Script*, because they sample articulation in a dense way; (c) the so-called "shibboleth sentence" from [6] (i.e., *She had your dark suit in greasy wash water all year.*), because it was repeated twice per speaking condition and because it includes strong examples of the corner vowels, relatively many consonants, and the troublesome phrase *wash water*; and, (d) isolated words *blow*, *blend*, *problem* and *row*, because each contains consonants and vowels that are strongly labial.

3 RESULTS

3.1 GENERAL OBSERVATIONS

In at least three broad ways, our talkers performed similarly when venting. One of these related to changes in positioning and overall movement of articulator markers. As a first example, we note that both talkers raised the jaw toward the maxillary teeth, to a position ~ 3mm higher than the average jaw position during normal speech. They also fixed the lips in a smile and held them back against the teeth, ~ 4-5 mm rearward of their normal average position. Postural changes for the lips and jaw were accompanied by large reductions in lip and jaw markers' respective movements. More than a century ago, Scripture [10] citing Flatau & Gutzmann [2] noted that "[i]n ventriloquism the movements of the lips and lower jaw are made as small as possible," and that "tongue articulations are greatly altered." We can quantify how much smaller, and how much altered, in terms of proportional changes in mid-sagittal areas enclosed by composite marker trajectories during the *Hunter Script*. Representative results are shown in Figure 1 for the female talker, who moved at least 50% more for all speech tasks than the male.

When venting, both talkers reduced mobility for lip (U and L) and jaw (M) markers by ~75-85%. Normal speech mobility areas shown in Figure 1 for U, L, and M markers

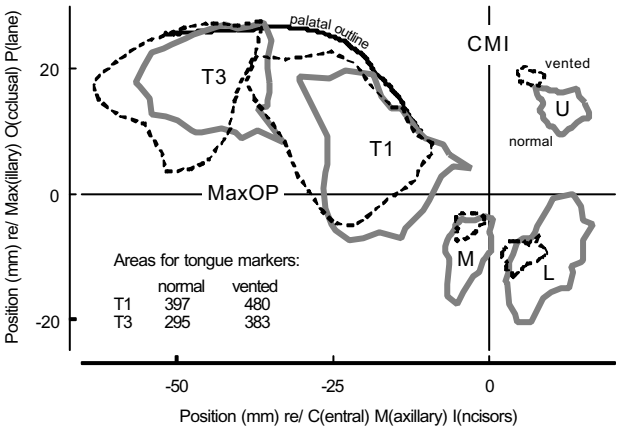

Figure 1: Marker mobility ranges from the *Hunter Script*

for the female are 41, 154, and 65 mm^2, while vented areas are 8, 24, and 13 mm^2, respectively. The male talker also reduced mobility of the three anterior tongue markers when venting (T1 by ~ 8%, and T2-3 by ~25%), but increased mobility of the dorsum marker (T4) by ~10%. In contrast, the female talker venting in her own and her main character's voices increased mobility of all tongue markers by at least ~ 20%. For shorter tasks than the *Hunter Script*, containing fewer sounds and words and less prosodic variation, changes in marker mobility could be much greater than those illustrated in Figure 1. As examples, during repetitions of the ~ 4-second shibboleth sentence recorded from the female talker, mobility areas for U, L and M markers were reduced to ~ 1, 5, and 2 mm^2, respectively, and mobility areas for tongue markers were increased by more than 90%, from an average area (computed across markers) of ~ 130 mm^2, to ~ 250mm^2.

A second broad way talkers performed similarly when venting was by speaking more slowly than normal. However, it is important to point out that venting-related timing adjustments were not uniform within or across tasks or talkers. Instead, most such adjustments seemed to be ad hoc responses to articulatory challenges imposed by sounds and sequences that are normally strongly labial. As an example, both talkers lengthened the *wash water* phrase in vented instances of the shibboleth sentence by ~ 300 ms (from an average duration of ~ 800 ms). The bilabial glides normally occurring as onsets of the successive strong words probably prompted them to do so, since each talker lengthened especially each word onset. The male seemed merely to stretch the acoustic and articulatory transitions into each (initial) nucleus by ~70-100 ms when venting, without making large modifications in the basic shapes of tongue-marker trajectories. He also "replaced" the normal medial 20-ms tap in *water* with a hyper-articulated 150-ms [t^h]. The female also extended

the onsets of both words by 125 ms or more when venting, separating the words by an "extra" silent interval of ~80 ms. But, her timing adjustments, unlike his, involved significant changes in tongue marker movements during the word-initial sounds. In each vented word, she seemed to "substitute" an exaggerated /l/-like sound for the normal initial /w/.

A third broad way talkers performed similarly when venting was by increasing average F0, and varying F0 more, presumably to create the impression of a younger, more animated talker. During the vented shibboleth sentence, the male (speaking in his character's voice) raised average F0 by ~ 30 Hz. The female speaking in her main character's voice did likewise, raising average F0 by ~ 40 Hz. Venting in her own voice for this task, she also raised average F0 relative to her normal voice but only by ~10 Hz.

3.2 SPECIAL LESSONS FROM *seed*

Both talkers repeated the isolated word *seed* relatively many times under each speaking condition. Articulatory and acoustic measures from multiple instances of this word provided estimates of intra-task reliability, and thus, bases for judging whether differences between paired single tasks produced under different speaking conditions were large enough to reflect condition-effects rather than mere production noise or measurement error. For both talkers, marker positions were measured at four acoustically-defined landmark events in each available replicate of *seed* (e.g., at frication onset and offset for the initial /s/, and oral closure and release for the final /d/), and standard distances were calculated by marker for each event. LPC and spectrographic methods were used to estimate {F1, F2, F3} frequencies at the time midpoint of the vowel /i/, and (F1,F2) standard distances were also calculated from these measures.

In terms of positional variability, the female talker was less variable overall than the male (median standard distances of 0.69 and 1.04 mm, respectively). Both talkers were less variable venting than speaking normally; less variable at frication and closure onsets than offsets; and, more variable for tongue than lip or jaw markers. In terms of acoustic variability, both talkers were comparably variable when venting with or without markers (median F1-F2 standard distances of ~50 Hz). The female was more variable than the male speaking normally, again with or without markers (standard distances of ~75 and ~25 Hz, respectively).

Close comparisons of trajectories traced by the forward tongue markers T1 and T2 during normal and vented *seed* revealed a significant speaking-condition effect for both talkers, though the difference between conditions was much more pronounced for the male. In general, both talkers held the tongue front higher in the mouth during vented than normal *seed*, and for the male, the shapes of the marker trajectories were radically altered. These result were contrary to our initial expectation that tongue movements during *seed* would not be much affected by venting. After all, its sounds involve mainly the positioning of the tongue relative to the palate, with no special role assigned to the lips or jaw.

Average (F1, F2) frequencies from replicates of *seed* revealed a main effect for the male talker that also seemed to generalize to vowels in other words (see below). On average, for the male, formant frequencies were higher in vented than normal *seed* (F1 by ~40 Hz, and F2 by ~200 Hz) spoken with or without markers in place. No comparable trend could be found in acoustic data for *seed* recorded from the female talker.

3.3 VOWELS

For the male talker, matched comparisons of F1 and F2 for normal and vented vowels produced in isolation, and in words {*she, greasy, had, all, wash, suit*} from the shibboleth sentence – measured in every case at about the time midpoint of the vowel – indicated a systematic effect of venting observed first among replicates of *seed*. Formant frequencies plotted in the lower left corner of Figure 2 for the male talker, for 5 of the 6 "shibboleth" words spoken normally 3 times, and vented twice, show again the trend for significantly raised F1 and F2 in his vented vowels.

Graphical comparisons of tongue, lip and jaw marker positions for vowels in three of these words (*she*, *had* and *suit*), measured at the same times as their formants, are plotted in the upper/right portions of the same figure.

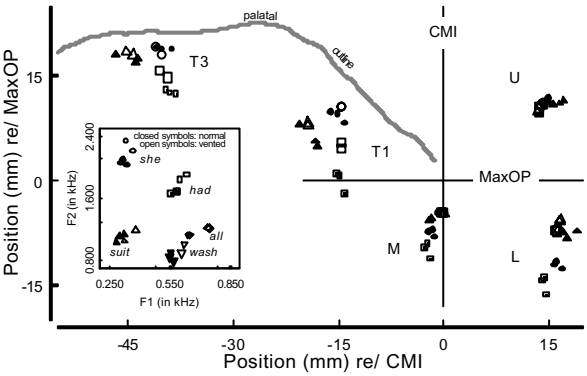

Figure 2: Formant frequencies and marker positions at midpoints of vowels in "shibboleth" words: male talker

What may be most noteworthy about these data is that the mouth positions of tongue markers **were about the same** for the male talker's vowels in *she* (or *suit*) spoken normally and vented. Tongue marker positions for the vowel in normal and vented *had* were also **at similar places along the vocal tract length**, though the vented positions are clearly raised toward the palate (interestingly, by about the same amount that the jaw marker was also raised toward the maxillary plane in vented examples of this word). Data from Figure 2, coupled with conclusions from [7] about the acoustical effects of jaw raising, suggest that acoustic differences between the male talker's vowels spoken normally and vented probably cannot be explained by differences in tongue position. Instead, the more likely explanation for formant frequency differences shown in Figure 2 rests with the retracted, spread lip configuration adopted by the

ISBN 1-876346-48-5

male talker when venting. What makes this example most interesting is the implication that this talker did **not** seem to modify tongue positions for vowels to compensate for acoustic perturbations induced by venting-related constraints on lip and jaw motion. Instead, he seemed to practice his craft in a notably different way from a male ventriloquist studied some 70 years ago by Huizinga [5], using radiography, and probably also from the female in our own study. Huizinga's male talker, and our female talker, both provided strong evidence for relatively large postural adjustments of the tongue during vented vowels. For our female talker, these adjustments seemed in fact to blur acoustic differences between vowels produced under the different speaking conditions.

3.4 SUBSTITUTIONS AND COMPENSATIONS

An outstanding feature of our female talker's vented speech was her inclusion of substitutions for certain consonants – what we take to be dental stops substituting for initial labial ones, and an /l/-like sound/configuration for initial /w/. She also produced easily-recognized compensatory postures and movements for rounded vowels, the labial off-glide in diphthongs, and for certain instances of the post-alveolar approximant /r/. The male talker did not behave in any comparable way. In Figure 3, we plot tongue marker trajectories for the word *blow*, spoken normally and vented, to illustrate two such adjustments. One clear difference between the words involves tongue configuration at word onset (where open circles indicate the tongue position characteristic of her substituted dental stop). The other clear difference involves the strongly retracted tongue position at word offset (represented by open triangles). Remarkably, the magnitude of the compensatory tongue retraction closely matched positional differences for lip markers measured at the same time.

4 CONCLUSION

By definition, vented speech is to be heard, not seen, and there are broadly different ways to accomplish the illusion. Strategies may range from relatively close approximation of normal articulatory targets during vented speech, constrained only by strong limits on lip and jaw motion, to relatively large-scale articulatory compensations and sound substitutions. Talkers who follow the first path – perhaps like the male in our study – seem to tolerate systematic acoustic differences between normal and vented speaking conditions at the expense of less variation in articulation. Venting talkers who follow the second path – perhaps like the female in our study – may adopt a contrasting strategy, compensating and substituting more to counter-act articulatory constraints, to reduce acoustic effects arising from such constraints.

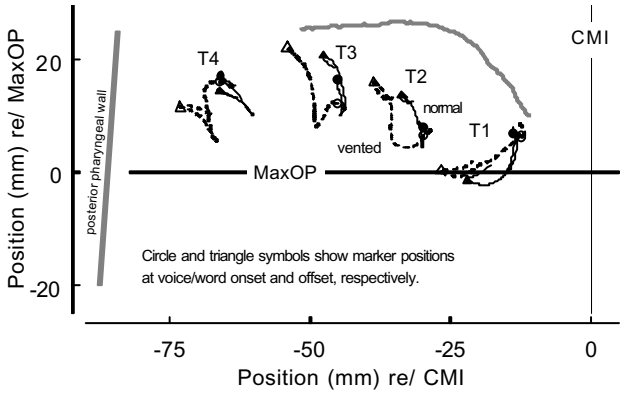

Figure 3: Marker trajectories for *blow*: female talker

ACKNOWLEDGMENTS

Research support was provided by USPHS/NIH Grant DC03723.

REFERENCES

[1] Crystal, T.H. and House, A.S. 1982. "Segmental durations in connected speech signals: Preliminary results," *Journal of the Acoustical Society of America*, 72, 705-716.

[2] Flatau, T.S. and Gutzmann, H. *Die Bauchrednerkunst*, Leipzig, 1894.

[3] Gay, T., Lindblom, B. and Lubker, J. 1981. "Production of bite-block vowels: Acoustic equivalence by selective compensation," *Journal of the Acoustical Society of America*, 69, 802-810.

[4] 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, 2854-2865.

[5] Huizinga, E. 1931. "Recherches sur un ventriloque Néerlandais," *Archives Néerlandaises de Phonetique Experimentale*, 6, 66-87.

[6] Lamel, L., Kassel, R. and Seneff, S. "Speech database development: Design and analysis of the acoustic-phonetic corpus," *Proceedings of the DARPA Speech Recognition Workshop*, 100-109, Palo Alto, CA: 1986.

[7] Lindblom, B. and Sundberg, J. 1971. "Acoustical consequences of lip, tongue, jaw, and larynx movement," *Journal of the Acoustical Society of America*, 50, 1166-1179.

[8] Maeda, S. 1991. "On articulatory and acoustic variabilities," *Journal of Phonetics*, 19, 321-331.

[9] Savariaux, C., Perrier, P. and Orliaguet, J.P. 1995. "Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube: A study of the control space in speech production," *Journal of the Acoustical Society of America*, 98, 2428-2442.

[10] Scripture, E.W. *The Elements of Experimental Phonetics*, New York: Charles Scribner's Sons, 1902.

[11] Westbury, J.R. *X-Ray Microbeam Speech ProductionDatabase User's Handbook*, Madison, WI: X-ray Microbeam Facility, 1994.

[12] Zuker R.M., Goldberg C.S. and Manktelow R.T. 2000. "Facial animation in children with Möbius syndrome after segmental gracilis muscle transplant," *Plastic & Reconstructive Surgery*, 106, 1-8.