Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland
Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola (Editors)

Postural Entrainment By Vocal Effort In Singing And Speech

Eric Vatikiotis-Bateson*1, Martin Oberg*2, Adriano Vilela Barbosa*3, Nancy Hermiston#4, Richard Kurth#5

*Linguistics, University of British Columbia, Vancouver, Canada
#School of Music, University of British Columbia, Vancouver, Canada

1evb@interchange.ubc.ca, 2bythebreeze@gmail.com, 3adriano.vilela@gmail.com,
4rkurth@interchange.ubc.ca, 5hermisto@interchange.ubc.ca

ABSTRACT

This study assesses the interaction of the postural control system and the production of expressive vocal behavior during speech and singing. In particular, we focus on the head, whose motions have been implicated for both postural control and spoken language production. How does head motion behavior simultaneously serve posture control and linguistic communication during vocalization? This study examines the interaction of these two subsystems by measuring the effects of different levels of vocal effort (loudness) on speech and singing. We show that as vocalizations becomes louder the correspondence between measures of head motion and speech acoustics become less complex and better coordinated spatiotemporally. In order to show that the head-voice coordination indeed concerns posture control, the same coordination effects are demonstrated for time-varying measures of body posture, measured with force plates under each of the performer's feet.

I. INTRODUCTION

That the head is an important component of postural control is well-established [1]. The head is responsible for vestibular and ocular contributions to balance. That the head plays perhaps multiple roles in the production and perception of language is also known [2-3]. We further know through countless observations that, without some degree of training, constraining head motion invariably reduces vocal amplitude. Finally, we know that both the postural control and speech production systems interact critically with the respiratory system.

If the head is simultaneously involved in two controlled time-varying tasks, it stands to reason that the two control regimes – one postural, the other linguistic – should interact [4]. However, the degree of interaction might be small and smoothly integrated when neither system is working very hard. In this study we examine how different levels of *vocal effort* – soft, normal, loud – influence measures of body posture and head motion during two vocal tasks: speaking and singing. We hypothesize that not only will the characteristics of head and body posture vary with vocal effort, but the interaction between the two systems will vary as well. We wish to examine this and two related predictions.

The preliminary results presented here suggest that vocal effort indeed affects posture (Figs. 2-3), and vocal effort effects on the coordination between the head and body (as measured at the feet) can be inferred from Figs. 2-3. The analysis of the instantaneous correlation between acoustic and postural measures across a range of temporal offsets (Fig. 4) suggests vocal effort has similar influences on both systems. Unfortunately, problems with calibrating the acoustics prevents more direct analysis of the acoustics.

A. Prediction 1

Because greater vocal effort marshals more physio-logical resources associated with breath control, the various components of the postural system – head, torso, legs – should become more highly coordinated spatially and temporally at greater levels of effort, and show some degree of interaction with the speech acoustics.

B. Prediction 2

When vocal effort is reduced, as in singing or speaking softly, components of the postural system should become less well-coordinated with the head.

II. METHODS

A. Subjects

Five undergraduates in the UBC School of Music Opera Program – three females, two males – participated in the study. All were in their 20's.

B. Materials

Subjects were asked to bring music that they liked, had memorized, and could sing without instrumental accompaniment.

C. Task

Subjects *sang* at three self-selected loudness levels: one termed *normal*, louder than normal, and softer than normal. Most subjects insisted on singing a different piece for the *loud* condition and at least one subject chose a third song for the *soft* condition. Subjects were also asked to recite their song(s) at three loudness levels. Interestingly, the first two subjects could not recite the texts of their music, so they and all subsequent subjects were told to *read* the texts.

D. Procedure

Subjects read or sang their pieces while standing on two Bertec force plates (one per foot) transducing 3D forces and moments (torques), and while wearing a lightweight head rig fitted with 6 infrared LEDs for transducing rigid body (6D) head motion. Voice recording was made via a Tram-50 lavalier micro-phone attached to the head rig approximately 20 cm above and behind the mouth. Force plate, head motion, and band-pass filtered (135 dB low-pass @ 2960 Hz) speech data were digitized via an OPTOTRAK 3020 system. The unfiltered speech signal and video of the subject (head-to-knees) were recorded to digital (Digital Beta) tape. Subjects repeated trials from one to three times according to their and/or the experimenters' satisfaction.

III. Results

In what follows, preliminary results are presented with the primary intent of describing the scope of the performance measures. The data are presented in several forms without the benefit of statistical tests for reliable contrasts, and without adjustment for the sizable variation in trial duration per condition. Due to the large number of measurements and performance conditions per subject (2 vocalization types X 3 vocal efforts), the overview presentations of the data in Sections III-B and III-D are necessarily less detailed than is ideal. An example of a more detailed descriptive presentation is given for one subject (*rf)* in Section III-C. Nevertheless, the mean trends do suggest smaller forces and less variable torques for *loud* than for *normal* effort conditions. *Soft* effort conditions are different from *normal* effort, but vary in their direction; sometimes showing larger and more variable forces than normal conditions.

A. Lower body posture

A first step in making sense of the different types of postural data is to examine the lower-body forces and torques that were recorded separately for the feet using a force plate under each foot. These measures, of course, are influenced by postural changes of the upper body (arms, torso, and head), but presumably record independent contributions of the lower body as well. Figure 1 shows the orientation of the force plate measures made for each foot; the x- and y- axes lie on the surface of the plate, while the z-axis is orthogonal to the plate along the gravitational axis.

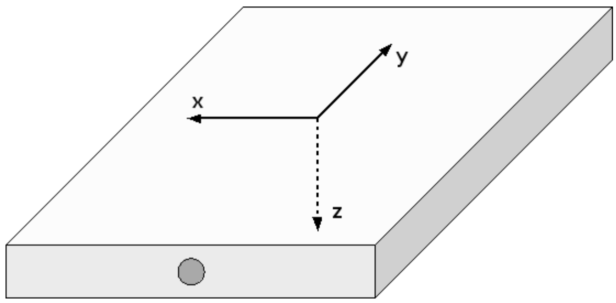

Fig.1. Schematic of 3D force plate measures. Forces were measured along each axis and moments (torques) around each axis. Subjects stood with one foot on each plate oriented along the long (y) axis, facing in the direction of the dot.

Figure 2 shows correspondences between the forces and torques of the two feet. The force correlation, shown in the top panel of the figure, confirms our observation that forces at the feet were nearly identical. No correlation was less than r = .80, and the vast majority were greater than r = .95, accounting for at least 90% of the variance. Torques, however, varied wildly within and across subjects, ranging over the course of a trial between nearly perfect correspondence (e.g., subject RF*,* trial #11) and nearly zero correspondence (e.g., subject GM, trials 10-11).

Thus, the linear forces exerted by the two feet are essentially the same at any given moment, even though the forces change somewhat through time (see below). Torques, on the other hand, show the rotational forces for the two feet to be highly variable and often uncoordinated over the time-course of a trial. A goal of subsequent analysis will be to determine if there is a more fine-grained pattern of coordination between the torques of the two-feet. This will be assessed using instantaneous correlation measures (for details, see [5]).

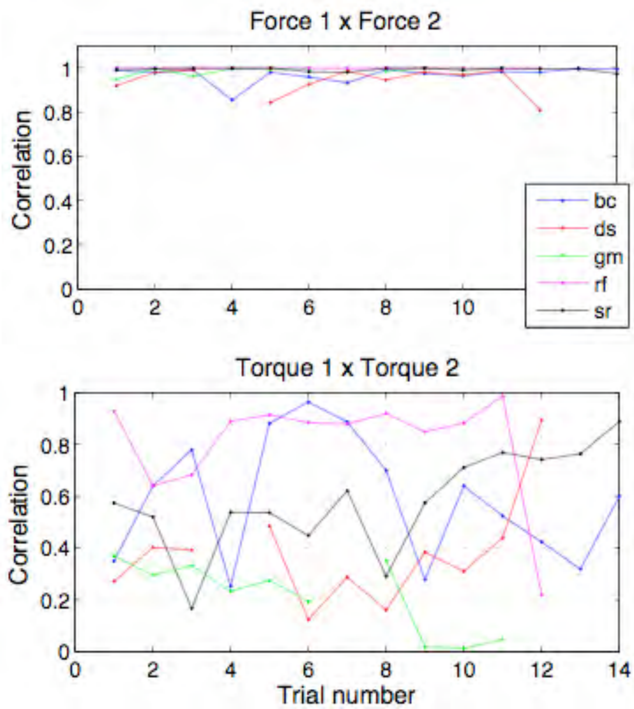

Fig. 2. Mean correlations for the forces (top) and torques (bottom) of the two force plates for all trials and all five subjects (legend).

B. Overview of descriptive results

Figure 3 presents mean results for the principal measures associated with rigid body head motion and foot forces. The panels are organized in vertical pairs, with one panel for Reading above and one for Singing below for each of four measures:

- 3D head translation computed as a *root mean square* (RMS);
- 3D force (RMS) for Force Plate 1 (right foot);
- 3D torque (RMS) for Force Plate 1;
- 3D torque (RMS) for Force Plate 2.

Since force was effectively the same for the two feet, force measures are given only for the right foot of each subject. In addition, head rotation was small and uncorrelated with any other measure or condition manipulation, so only the linear translations of the head are plotted here. Empty columns signify no usable data for that condition.

While the within subject results would be easier to decipher in subject-specific box plots, as exemplified for Subject SR in Fig. 4, bar plots readily show the large variability within conditions and across subjects. All measures vary considerably across subjects. More disturbing, however, is the within-subject variability between vocalization and effort conditions, statistical analysis is not needed to verify the absence of clear trends in the means; the large standard deviations (and small differences in means) are sufficient. There is a hint that Reading and Singing may differ, but the means alone are insufficient to show the difference.

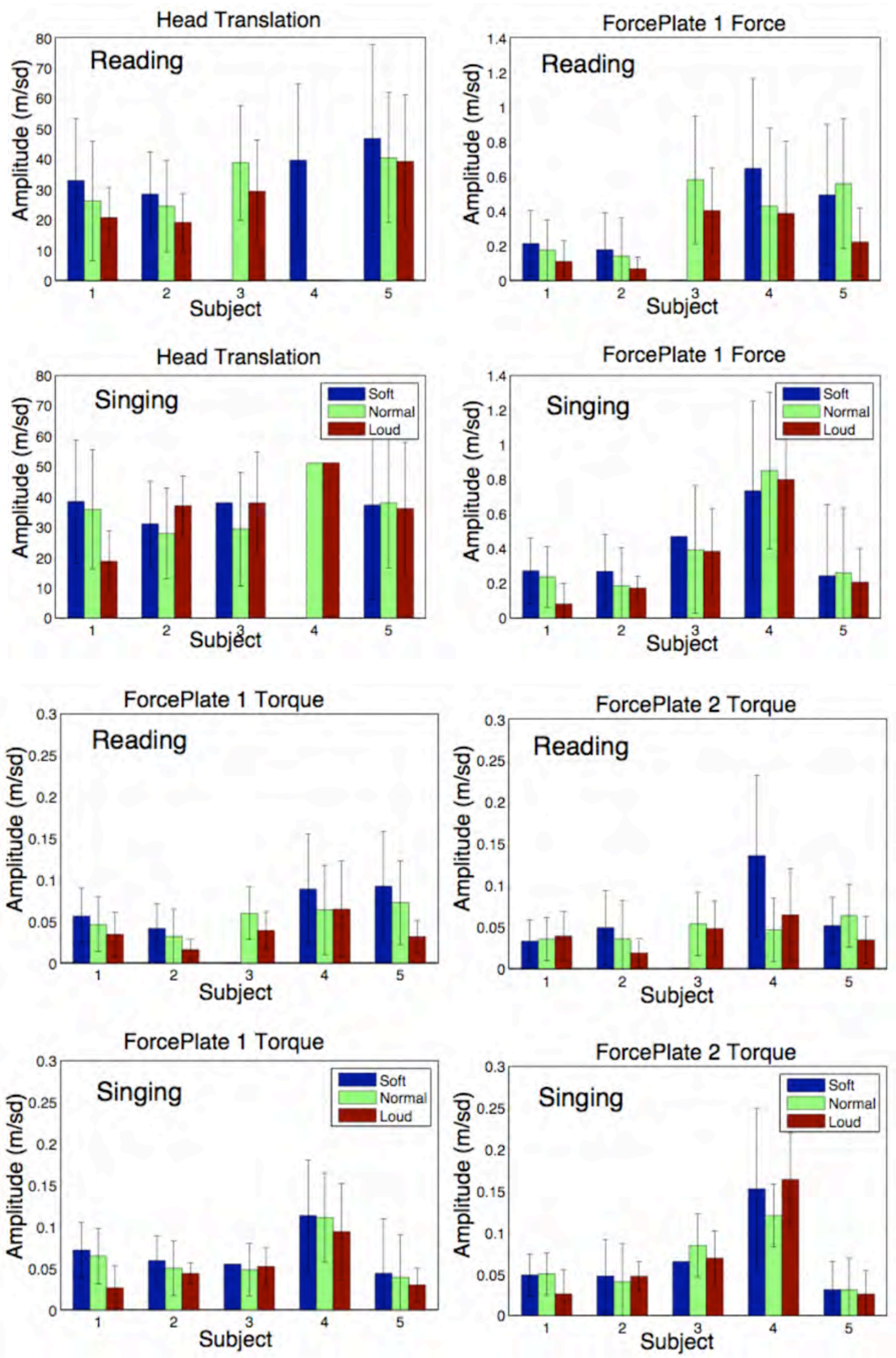

Fig. 3. RMS amplitude means and standard deviations (error bars) plotted by measurement type, vocalization condition, subject, and vocal effort. Force is shown for right foot only (see text, Section III-A. and also Fig. 2).

C. Descriptive results for one subject

Figure 4 shows almost the same results for subject SR in box plot form as shown in Figure 3. The only difference is that the RMS for head motion is computed using the three translations along *and* the three rotations around the three coordinate axes. This view affords greater optimism that there may be differences due to vocalization type and vocal effort condition than suggested by the bar plots. For example, there are more large values (e.g., head translation) associated with *reading* than *singing*. Comparing normal and loud productions during reading, her amplitudes for the loud condition are generally the same as for the normal condition, but the variability appears to substantially higher for the loud condition in three cases. This is what we would expect. However, all of this subject's values for *loud* singing, including her acoustic amplitude values, are suspiciously small. This appears to be the case for most of the subjects and we believe points up a flaw in our method, in that we allowed singers to sing different songs in the loud conditions when they claimed they could not sing the originally chosen song any louder. This will be remedied in a follow-up study using more carefully selected materials that allow us to enforce distinct levels of vocal effort.

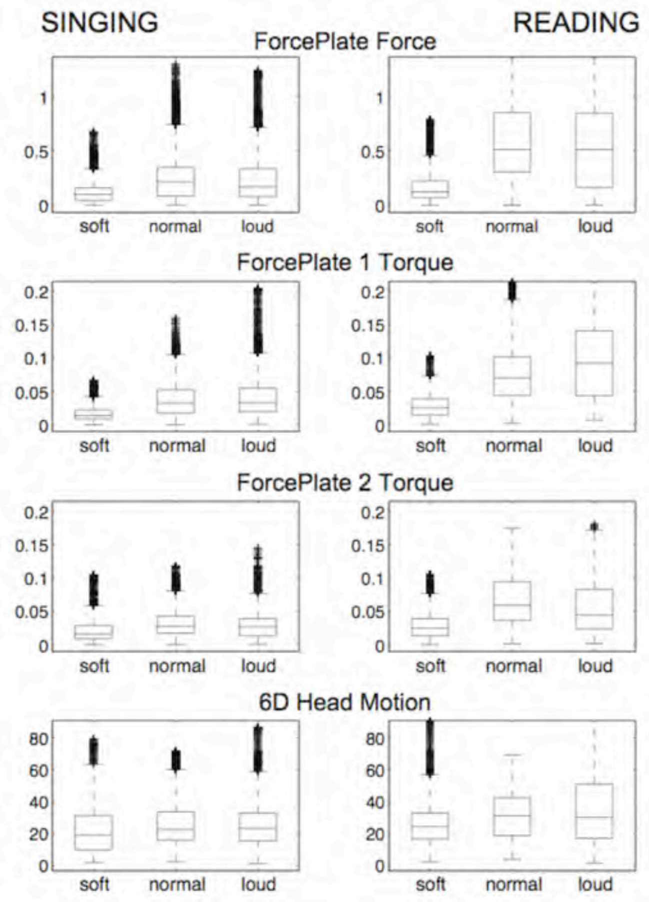

Fig. 4. Box plots of RMS values for principal physical measures for subject SR show median (horizontal line within box) and distribution of data at specific offsets from the median; whiskers show individual values more than 1.5 times the distance from the median to the relevant edge of the bounding box (25% of the data).

D. Instantaneous correlation analysis

Due to an error in the data recording, the acoustic amplitude cannot be calibrated for 4 of the 5 subjects; subject SR (described above) being the exception. As a result, the relations between physical behavior and acoustic correlates of vocal effort must be examined in ways independent of absolute values of amplitude. To achieve this, an algorithm we developed [5] that computes the instantaneous correlation between two signals is used to compare the time-course of acoustic amplitude to the time-varying behavior of each of the physical measures we made. A second feature of the algorithm is that signals can be compared across a user-selected range of temporal offsets. This is particularly important when comparing signals whose closest correspondence might not be at zero phase-offset, or more interestingly even when the phase of highest correlation between the two signals might vary through time.

In order to assess the correspondences between physical and acoustic signals, the complex histograms shown in Figure 5 were computed. What are shown on the y-axis are the instances where the instantaneous correspondence exceeded $r = .5$ (chosen because 25% of the variance is too large to be coincidental) for the range of temporal offsets 6 seconds before and after zero-offset (x-axis).

Several things should be noticed immediately. First, there was a substantial number of instances (*hits*) where the instantaneous correlation exceeded $r=.5$ for every condition of vocal effort and vocalization type. Second, there is an obvious difference between singing and reading as shown by the density and definition of the histogram patterns. The number, thickness, and temporal spacing of the spikes in the histograms all point to a fundamental coordination difference between singing and reading. The high density of hits for singing suggests a higher degree of and more continuous coordination between the acoustic and the postural measures. The reading condition, on the other hand, shows high incidence of coordination at distinct, more narrowly defined, temporal offsets. In both cases, the temporal separation (x-axis) between the spikes – approximately 2 seconds for reading and 2.5 seconds for singing – indicates rhythmic pattern consistency. For example, a signal's behavior during one syllable is not only correlated with another signal's behavior during the same syllable, but also with that second signal's behavior in syllables preceding and following it. Third, a similar, albeit less distinct, difference may hold for the effort conditions. Specifically, loud conditions display more sparse patterns of correspondence, with loud reading being more sparse than loud singing.

This last finding in particular suggests that coordination may be reduced at higher levels of vocal effort, which goes against our initial prediction that as effort level increases, coordination should increase. On the other hand, the more sparse, narrowly prescribed patterns of correspondence shown by both loud singing and reading may indicate a change in the type of coordination amenable to reduced coordination overall, but more precise instances of coordination. More detailed analysis is currently underway that should help us flesh out this story.

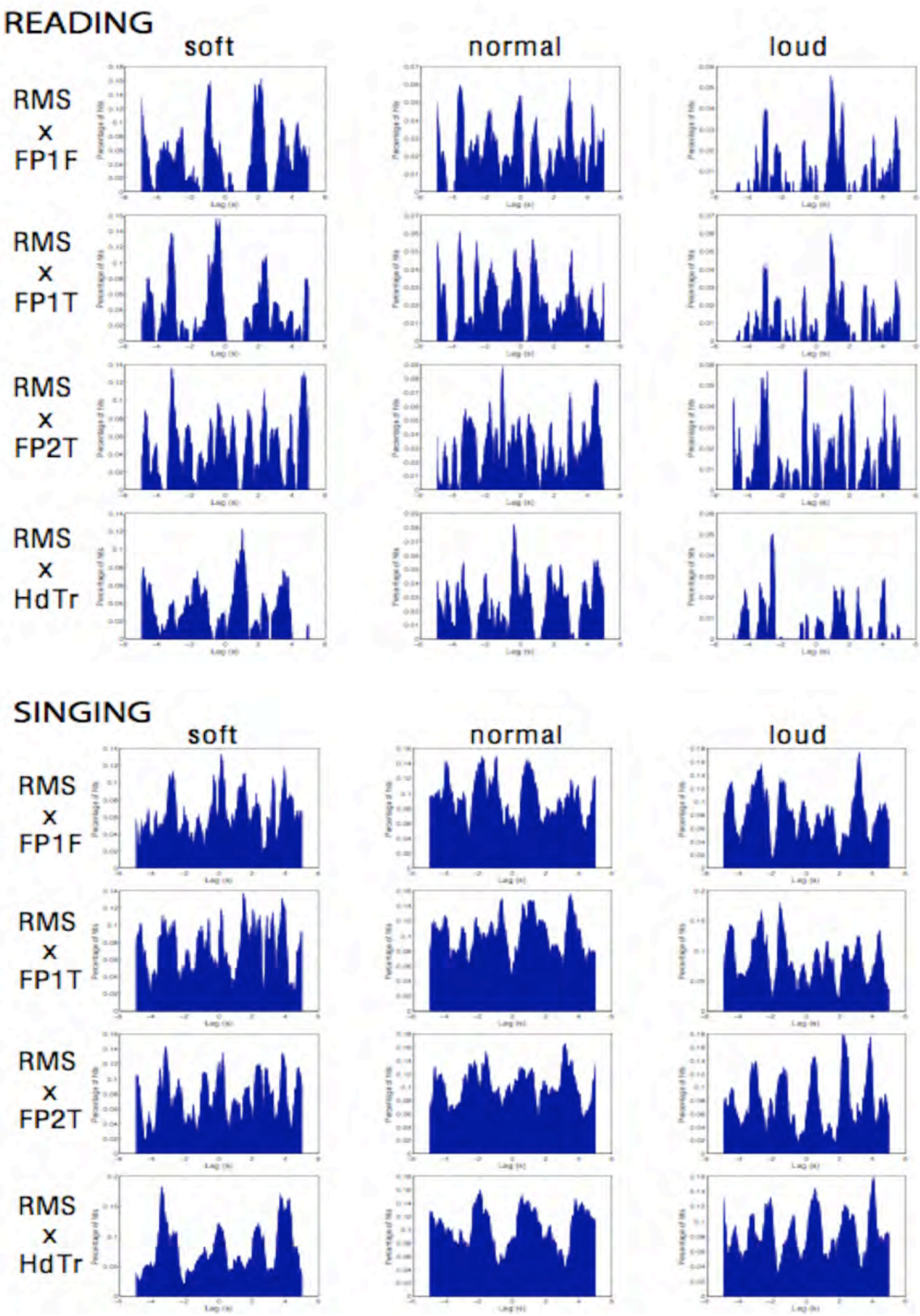

Fig. 5. Histograms showing the proportion of *hits* (y-axis) over the course of a trial when the instantaneous correlation reached a threshold of r=0.5 are computed for a range of temporal offsets (x-axis). For both reading (top) and singing (bottom), RMS is compared to RMS values for force on one plate (FP1F), torque on both plates (FP1T, FP2T), and head translation (HdTr).

IV. GENERAL DISCUSSION

It remains to be seen whether or not the data of this study show reliable magnitude differences for different levels of vocal effort. Indeed, we intend to re-run the study using more stringently selected and controlled materials for singing and speaking. The current study was not intended to have a reading condition. This arose when we discovered that our singers could not recite from memory the lyrics when they were being spoken instead of sung.

Nevertheless, it is clear that there *are* performance differences in the measures of head motion, posture, and the acoustics. We learned that several measures probably need not concern us in future analysis; notably head roatation, since the coordinated head motion behavior is almost entirely translational, rather than rotational. Additionally, we learned that the linear forces are the same for the two feet. Unless the performer is hopping from foot to foot, this is likely to be the case for any fixed stance performance.

We further learned that clear correspondences exist between the postural and vocalization associated with vocal production, and that these correspondences vary with vocal effort. This was shown by computing the instantaneous correlation between various signal pairs across a range of temporal offsets using an algorithm devised expressly to assess coordination between time-varying measures. Other recent applications of the algorithm include the coordination of speech and visible gestures of the head and hands [6], and the coordination between a performer and a large audience, which itself can be assessed for internal synchronization [7].

Despite the evidence that coordination between the vocal and postural systems interacts with vocal effort, we do not yet have anything that speaks directly to our initial predictions that coordination between the postural control system and vocalization becomes more critical at higher levels of vocal effort. There are myriad reasons why this might be so including the possibility that we have not yet formulated the right question. Another possibility is that it is a mistake to use young opera singers-in-training as subjects. Our opera students are trained from the start to protect their voices, hence their preference to sing a "louder" song rather than sing a song louder. Also, because both operatic singing and reading are highly stylized, differences in coordination due to vocal register may be masked. Furthermore, reading text from a quarto-sized book of music held in front of the body introduces another postural dimension that may suppress the coordination that would be seen when vocalizing with one's hands free.

Finally, lack of calibration for acoustic amplitude limits our options for cross-domain measurement comparison. Thus, we will re-run the experiment with singers who can sing and recite the same songs at different loudness levels. We will also record spoken language samples, both spontaneous and scripted, using non-singers who may be willing to modulate their amplitude more than the trained singers, and compare their productions with the spoken trials of the singers.

ACKNOWLEDGMENT

Funding for this research was provided by the National Science and Engineering Research Council (Discovery Grant to Eric Vatikiotis-Bateson and New Media Initiative Grant to Sidney Fels) and the Canada Foundation for Innovation.

REFERENCES

1 Dault, M. C., Yardley, L., & Frank, J. S. (2003). Does Articulation Contribute to Modifications of Postural Control During Dual-Task Paradigms? Cognitive Brain Research, 16(3), 434--440.

2 Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science, 15*(2), 133-137.

3 Yehia, H. C., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion, and speech acoustics. *Journal of Phonetics, 30*(3), 555-568.

4 Rimmer, K. P., Ford, G. T., & Whitelaw, W. A. (1995). Interaction Between Postural and Respiratory Control of Human Intercostal Muscles Journal of Applied Physiology, 79, 1556-1561.

5 Barbosa, A.V., H.C. Yehia, and E. Vatikiotis-Bateson, *Temporal characterization of auditory-visual coupling of speech*, in *Proceeding of Meetings on Acoustics*, A.D. Pierce, Editor. 2008, American Institute of Physics; p. EL1-13.

6 Barbosa, A.V., H.C. Yehia, and E. Vatikiotis-Bateson, *Linguistically Valid Movement Behavior Measured Non-Invasively*, in *Auditory and Visual Speech Processing -- AVSP08*, R. Goecke, P. Lucey, and S. Lucey, Editors. 2008, Causal Productions: Moreton Island, Australia. p. 173-177.

7 Barbosa, A. V., Munhall, K. G., Yehia, H. C., & Vatikiotis-Bateson, E. (2009). Quantifying the Time-Varying Coordination Between Performer and Audience. Retrieved Acoustical Society of America 157th Meeting Lay Language Papers, World Wide Press Room (http://www.acoustics.org/press/157th/bateson.html)