Investigating embodied pitch cognition in improvising saxophonists: Discussion and an experimental design

Introduction

There is no life without movement: from the accumulation and firing of electrical signals within our neural circuit and nervous systems to the interaction of atoms in the surrounding environment with our own bodies, movement is a prerequisite for life. The structural concepts that an organism builds, uses to perceive, understand and navigate its moving environment depend on the constitution of its body – that is, that the nature of an organism's body will determine the extent and quality of it's perception. But the nature of how these structural concepts are represented is still under debate. Cognitive science predominantly adheres to a computational theory of mind in which "mental processes proceed algorithmically, operating on symbolic representations" and have "determinate starting and ending points" which "limit its investigations to processes within the head, without regard for the world outside the organism" (Shapiro 2011). However, ideomotor theory suggests that actions are represented in terms of their perceivable effects, and that the world inside is functionally equivalent to the moving world as perceived outside in so much as it is coded "in a common representational medium by feature codes with distal reference" (Hommel et al 2001).

In order to test this, research has been conducted looking into whether perceptual stimuli have a priming effect on actions, and some of this research has been musically based. In an experiment in which action-effect (A-E) associations were investigated by having pianists and guitarists respond on their instruments to visual stimuli with auditory distractors, Drost et al (2007) found a significant interference effect amongst the incongruent and congruent condition difference only when the timbral quality of the distractor matched that of the instrument played by the participant, therefore showing that "integrated A-E associations primarily seem to consist of a specific component on a sensory-motor level involving the own instrument" (Drost et al 2007), thereby supporting ideomotor theory. Murphy replicated this study successfully for violinists in her Cambridge undergraduate submission for Perception and Performance (Murphy, L undergraduate report).

As above, ideomotor theory suggests that actions are represented in terms of their perceivable effects and therefore that perceiving and action planning are functionally equivalent. However two questions emerge from the above. Whilst many studies have looked into whether perceptual stimuli have a priming effect on action, ideomotor theory states that it is entirely possible for the opposite to be true, and that "the ideomotor approach does not deny the opposite direction of influence on perception while a certain action is being performed" (Shin et al 2010), comparatively few studies have been conducted looking into this inverse, namely whether actions, or representations of actions, prime a perceptual expectation leading to changes in response time or heightened accuracy, particularly within the musical fields. And secondly, as also seen, many of the existing studies look into the performers of instruments demanding a certain threshold of physicality in order to play: pianists, violinists and guitarists mainly use their arms, hands and fingers to produce the sound on their instruments. Although to a large extent posture and frame control are important aspects of consistent performance on all these instruments, there are other common instruments that demand control over additional variables: with respect to the diaphragm and breath control, larynx and tongue, jaw and facial muscles, wind and brass instruments in general require more physical processes in order to produce consistent sound.

This prompts some interesting questions to be asked. If, as in Drost et al's conclusion, A-E associations exist on a sensory-motor level involving the performer's own instrument, do instruments requiring more coordinated physical processes produce a greater quantity, or stronger quality, of integrated A-E associations? If the main focus of research to date has predominantly looked at digit and lower arm movements in pianists, guitarists and violinists, we can also ask to what extent these sensory effects permeate core movements (as opposed to digit lower arm) or physiological ones (such as breathing or breath control) that are involved in the production of sounds on wind and brass instruments? And does the strength of the A-E association vary across musicians who, all other things being equal, perform music that requires a potentially different process of recall, such as musicians who consider themselves to be improvisers as opposed to those who primarily perform written music?

In an effort to begin to address some of these questions, a within-group study was devised to investigate whether there was an effect of three breathing conditions (exhaling, inhaling, held breath) upon response time and error rate in saxophonists who considered themselves improvisers. On the saxophone, sound is produced by maintaining a consistent pressurized airflow from the diaphragm, is maintained and directed by an embouchure (a combination of facial muscles and lips to maintain pressure through the mouthpiece), from where the air flows into the instrument and is mapped into pitches by pressing combinations of keys. The aim was to find out whether training on the saxophone leads to A-E associations relating to a the specific breathing conditions employed in the production of sound, and therefore a quickened response time and lessened error rate when these are replicated. Only improvisers were chosen for this study, with a plan to compare results in the future to non-improvisers under the same conditions. Response times and error rates were measured to determine whether there was any interference from the breathing conditions. There were two hypotheses: 1) That reaction times would be slower and error rate higher in the inhaling breathing condition than in the exhaling and held breath conditions; and 2) that reaction times would be slower for incongruent stimuli than for congruent stimuli across conditions (as in Drost et al, and in Murphy, L undergraduate report).

This was not the only methodological design considered: initially designs focused around designing and writing a python script to monitor breathing rates through a midi controller. However it was decided not to monitor breathing characteristics as it would have proven difficult to obtain a reliable and realistic controller: an EWI (electronic wind instrument) was considered as a controller, but the touch sensors were tested and deemed too variable to capture response time and error data accurately and reliably. There was also a second controller considered, named WX5, but it transpired that this model was not legal to be sold in the UK under EU regulations, and so this was ruled out as an option.

Method

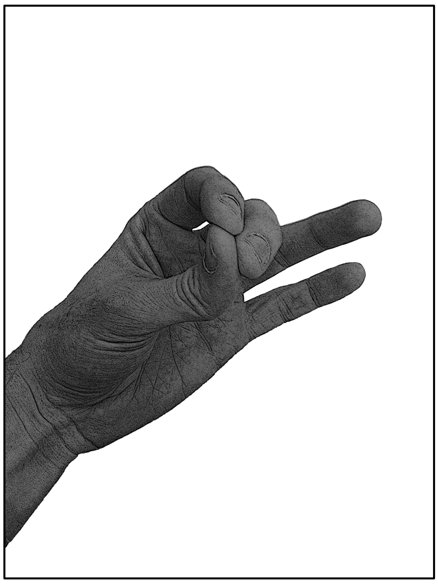
Participants

A total of 14 experienced saxophonists (all male; 3 left handed, all other right handed; 5 with corrected eyesight, all other with normal eyesight; 5 tenor saxophonists, all other alto saxophonists; 10 undergraduate, 1 graduate, 1 postgraduate, 2 professional; age in years: M = 22.9, SD = 4; years of playing: M = 10.7, SD = 4.3) participated in the experiment. None of the subjects had perfect pitch and none of them had hearing problems or corrections. When asked whether they considered themselves improvisers, all subjects responded 3 or higher on a scale from 1-5 where 1 = not an improviser, 5 = improviser. No monetary payment was offered for participating in the experiment, although subjects were offered sweets as a reward for taking part.

Material and equipment

Participants responded using the keyboard of a MacBook Pro computer. The keys used to respond in the experiment were *C* and *M*. The experiment was conducted in a small, well-lit, isolated room. Participants were seated facing the screen at a comfortable distance before the computer, with their left hand index finger over the *C* key and their right hand index finger over the *M* key. Visual stimuli were presented on the computer screen, which was positioned on a desk at a comfortable height in front of the candidate. Auditory stimuli were played using Sony Studio Monitor over-ear headphones. Superlab software was both used to control the presentation of the stimuli and to record participant responses to the trials for analysis.

The four visual stimuli consisted of a black and white image of a hand simulating the fingerings for a lower register *g*, *a*, middle *b* and middle *c* on the saxophone, on a white background and no border (13.5x10cm) (see fig. 1). A total of 10 auditory stimuli were used for each experiment. Trials with adjusted and transposed auditory stimuli were used for alto and tenor saxophonists. They consisted of the notes *g*, *a*, *b*, *c*, and all other chromatic notes from a whole tone above the highest note (*d* above saxophone middle *c*) to a whole tone below the lowest note (low *f* on the saxophone). All visual and auditory stimuli were recorded and made by the experimenter (male, age = 27 years, 16 years training on the saxophone). The auditory stimuli were recorded from real instruments on a Zoom H4n handheld recorder and the samples were checked and edited for consistent onset, offset, tuning and intonation. The duration of each auditory stimulus was 1,000 ms.



(above) Figure 1. Scored notation showing the four congruent sounding pitches (bar 1), the six other notes used in the study (bar 2), and the range of the pitches used in the study (bar 3).

(left) Figure 2. An example of one of the four images presented to participants, in conjunction with the auditory stimuli. This fingering representation is of an *a* fingering, and from this it can be seen that both index and middle fingers are depressed, whereas the ring and little fingers are relaxed.

Procedure

In the first block a black fixation cross on white background (0.6 cm diameter) was shown for 1,500 ms at the centre of the screen. This was followed by the presentation of the visual stimulus, which remained on screen until the end of the trial. After 500 ms of the image appearing, the auditory stimulus (either congruent or incongruent with the presented image) was played for 1,000 ms. Participants were instructed to use the *C* and *M* keys to respond as to whether the visual and auditory stimuli were congruent or incongruent. Upon a response from the participant, or following a time-out after 2,500 ms, this screen was followed by the word ‘next’, in black on a white background for 1,250 ms before beginning the next trial. In subsequent blocks the initial fixation cross was replaced, for the same duration, with the word ‘exhale’ in block 2, ‘inhale’ in block 3, and ‘hold breath’ in block 4 to cue breathing conditions.

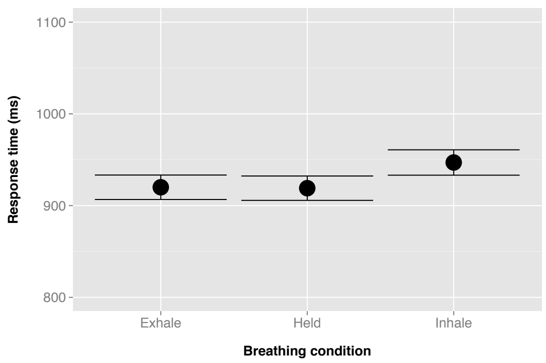
In every block trials were presented as either congruent or incongruent with the visual stimuli for *g*, *a*, *b* or *c* saxophone fingerings. The incongruent tones were selected from the 9 tones in the note pool that mismatched the visual stimulus (if the visual stimulus was of a *g* fingering, the notes *f*, *f#*, *g#*, *a*, *a#*, *b*, *c*, *c#*, *d* were used as the incongruent note pool). Each of these incongruent notes was selected once per sequence, but the order of each sequence was generated randomly for every block. There was a random sequence for each of the four mismatched images (36 trials), and these were randomly mixed with 9 trials for each congruent visual-audio pairing (36 trials) per block.

The breathing instructions were introduced at the start of the relevant block. For the 'exhale' condition, the participants were asked to exhale lightly through their mouths from the visual cue till the time they responded to the trial. Similarly for the 'inhale' condition, participants were asked to inhale lightly through their mouths from the visual cue through to their response. For the 'held breath' condition, the participants were asked not to inhale sharply and hold, but to halt their breathing at the point in their breathing cycle that the visual cue was presented until they responded the stimuli. Breathing rates, peak flow, breath onset and offset were not measured.

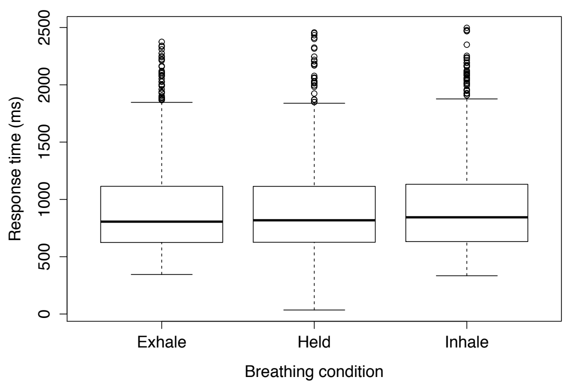
The first four practice trials in the experiment gave participants feedback on their choices after responding (correct or incorrect) and introduced all of the visual stimuli that were to be used in the forthcoming experiment. Neither the main practice block or subsequent blocks offered written or visual feedback, nor there was any auditory feedback throughout the experiment. These four trials were followed by a short break for the participant to ask questions before resuming. Participants received an average of 62 practice trials in block 1 before moving on to the subsequent blocks, and were not made aware of any breathing conditions until the beginning of the relevant block. As above, blocks 2, 3 and 4 consisted of 72 trials each, with each block containing an equal amount of randomly ordered congruences and incongruences. After each block, the participant was offered a short break in which the instructions for breathing were given for the forthcoming block. Practice block 1 was always the first to take place, but the order in which blocks 2, 3 and 4 were presented was rotated between participants in an attempt to counterbalance any learning effect (1234, 1243, 1324, 1342, 1423, 1432), and the response keys were reversed for each of the six rotations to counteract any handedness bias (A1234, B1234, A1243, B1243 etc.). Additionally, in an attempt to neutralize any confounding memory-aiding mnemonic, the variables were presented to the participant as 'congruent/incongruent' where *C* was the congruent response key, and 'matching/not matching' where *M* was the congruent response key. Each participant was presented with written instructions detailing the nature of the first block, which were then discussed and clarified in a consistent manner by the experimenter. They were also told that the experimenter was looking for their fastest reaction time and to answer as quickly and accurately as possible to each trial. After each briefing they were reminded of the key orientation for responses. Consent forms were obtained from each participant, and each session lasted about 50 minutes.

Results

Trials in which no response was made were removed from the data prior to analysis. Additionally, one of the participants was unable to correctly follow the experimental instructions and was also removed from the data. Participants' mean response times per condition, and error rate for congruent and incongruent stimuli, are given in *Figure 6* (original participant numbers have been changed).



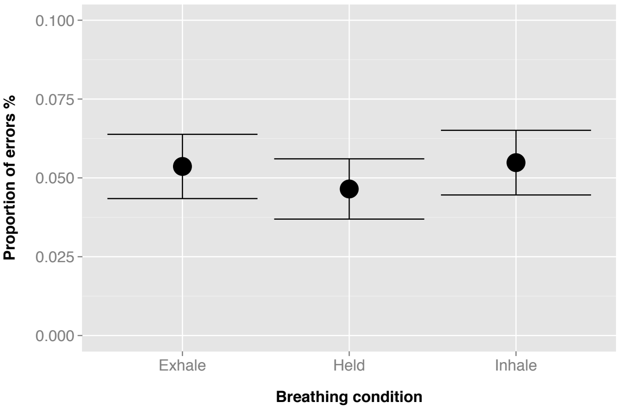
(above) Figure 3. Average response time by breathing condition with exhale, held and inhale (left to right, x axis) against response time (ms, y axis).



(above) Figure 4. Boxplot of response times per condition with exhale, held and inhale (left to right, x axis) against response time (ms, y axis).

Response times (both normal and log transformed) and error rates were submitted to separate one way repeated-measures ANOVAs with breath (exhale, inhale, held) as a factor. This analysis revealed that there were no significant effects of breath upon either response time measures: response time ANOVA ($F(2,11) = 1.165$, $p = 0.347$); log transformed response time ANOVA ($F(2,11) = 0.818$, $p = 0.467$). However, there was a marginally significant effect of breath upon error rates ($F(2,11) = 3.394$, $p = 0.071$). Follow up comparisons indicated that this partial significance was due to less errors occurring in the held breath condition compared with both of the exhalation and inhalation conditions. This trend did reach significance without corrections ($p = 0.05$); however it did not when the Bonferroni correction was used ($p = 0.162$).

A post-hoc analysis showed that response times across breathing conditions were consistently faster when reacting to incongruent stimuli than to congruent, and that errors occurred only for incongruent stimuli (see *Figure 6*).



(left) Figure 5. Proportion of errors by breathing condition, with Exhale, Held and Inhale (left to right, x axis) against proportion of errors (%, y axis).

	Response Times								
	B2		B3		B4			percentage errors	
	congruent	incongruent	congruent	incongruent	congruent	incongruent	total errors	congruent	incongruent
p1	1403.83	1099.48	1344.53	1272.43	1183.47	946.86	23	0%	100%
p2	1357.55	948.48	1432.86	1089.00	1444.06	914.93	36	0%	100%
p3	1270.42	870.72	1198.85	1037.68	1357.24	1038.29	14	0%	100%
p4	617.61	544.24	524.56	515.54	545.97	499.08	45	0%	100%
p5	793.69	632.03	679.97	581.50	621.81	515.63	23	0%	100%
p6	997.22	840.90	1002.19	831.15	1057.78	902.44	53	0%	100%
p7	1198.54	1146.86	1202.75	1085.06	995.15	962.49	8	0%	100%
p8	776.11	659.15	844.39	690.08	701.50	667.77	39	0%	100%
p9	944.03	925.03	848.25	839.84	880.67	930.18	15	0%	100%
p10	672.53	610.57	762.11	757.62	679.39	707.13	53	0%	100%
p11	1047.00	818.82	1112.67	785.31	1166.94	823.28	4	0%	100%
p12	580.69	604.61	656.86	670.80	882.14	722.96	42	0%	100%
p13	1181.47	1020.21	1353.83	1133.87	1354.43	1268.06	11	0%	100%
TT	917.19	765.79	925.99	806.42	919.32	778.51	366	0%	100%

(above) Figure 6. Table showing participants' mean response times across conditions, with the mean across all participants displayed at the bottom and error proportions in the far right. Original participant numbers have been changed.

Discussion

Effect of breathing conditions on response times and error rates

The first aim of this pilot experiment was to determine whether reaction times would be affected by breathing conditions, and whether breathing conditions consistent with playing the saxophone (exhale, held breath) would prime faster response times. In individual instances the results for response times did show trends towards supporting the hypothesis –

however as a whole the results did not support the hypothesis and contained a great degree of variance, and therefore no conclusive finding relating the breathing condition to response time was made. A confounding factor in this could have been the time some participants took in order to respond: in some cases the response took between 1,250 and 2,500 ms, and it could be said that the timing data collected in these instances was not a reaction time, and that timings over a certain threshold could be interpreted as being a considered response, which in turn may affect the types of processes occurring and not necessarily represent any reactive ideomotoric effects.

Another aim of the first hypothesis was to determine whether error rates in responses were affected by breathing conditions, and whether breathing conditions that were consistent with playing the saxophone, as above, would return a lower rate of errors in the responses. Here a marginally significant effect of breath upon error rate was seen, with the held breath condition returning a lower rate of errors across participants than the exhale and inhale conditions, which supports the hypothesis. However, as was the case for response times, there was a high degree of variance and more participants and data would have been required from in order to draw any meaningful conclusions. The fact that there is some evidence here to support the hypothesis suggests that it would be worth looking into conducting further studies on a wider range of participants to determine whether the combination of more data and a better design regarding the slow reaction response rate would clarify the results and show some more significant differences.

Response times of congruent vs incongruent

The second aim of this experiment was to investigate whether responses to congruent stimuli would be faster when compared with those to incongruent stimuli. The design here was different to other studies which had looked at response times to auditory stimuli on actual instruments: but although here the participants were being primed by a visual representation of a fingering and were not responding on their instruments, a faster response time for congruent conditions would have seemed to be in line with findings made in various other studies in this area, and was expected. In fact what occurred was quite the opposite: response times for incongruent stimuli were faster in both the exhale, inhale and held breath conditions, as seen in *Figure 6*. Although unexpected, this may suggest that the processes used to determine the congruity of a stimulus in the manner presented in this study are subject to a kind of sensory feedback mechanism which, although "directed by [the] anticipatory representation of its own feedback or of feedback from the reaction to a goal to which the response leads", discards incongruent pairs of stimuli quicker than it affirms congruent ones (Greenwold 1970). Again, given the variance amongst the results there should be no definitive conclusions drawn here, but this certainly highlights an interesting area for further investigation arising from this brief pilot study.

Bibliography

Drost, Ulrich; Rieger, Martina; Prinz, Wolfgang, “Instrument specificity in experienced musicians” (The Quarterly Journal of Experimental Psychology, 2007, 60-4, 527-533)

Greenwold, Anthony, “Sensory feedback mechanisms in performance control” (Psychological Review, 1970, Vol 77, No. 2, 73-99)

Hommel, Bernhard; Musseler, Jochen; Aschersleben, Gisa; Prinz, Wolfgang, “The Theory of Event Coding (TEC): A framework for perception and action planning” (Behavioural and Brain Sciences, 2001, 24, 849-937)

Murphy, Lydia, “Does what we hear affect what we play? Investigating the effects of pitch and timbre on bowing in violinists” (Cambridge University: unpublished undergraduate report)

Shapiro, Lawrence, *Embodied Cognition* (Abingdon: Routledge, 2011)

Shin, Yun Kyoung; Proctor, Robert; Capaldi, E, “A review of contemporary ideomotor theory” (Psychological Bulletin, 2010, Vol 136, No. 6, 943-974)