E-mic: Extended Mic-stand Interface Controller

Donna Hewitt
University of Western Sydney
Locked Bag 1797, Penrith South DC
NSW Australia 1797
+61 2 9300 8485
dhewitt@bigpond.net.au

Ian Stevenson
Audile Design
2/4 Peckham Ave, Chatswood
NSW Australia 2067
+61 (0)407135475
audile@bigpond.com

ABSTRACT
This paper describes work in progress for the development of a gestural controller interface for contemporary vocal performance and electronic processing. The paper includes a preliminary investigation of the gestures and movements of vocalists who use microphones and microphone stands. This repertoire of gestures forms the foundation of a well-practiced 'language' and social code for communication between performers and audiences and serves as a basis for alternate controller design principles. A prototype design, based on a modified microphone stand, is presented along with a discussion of possible controller mapping strategies and identification of directions for future research.

Keywords
Alternate controller, gesture, microphone technique, vocal performance, performance interface, electronic music.

1. INTRODUCTION
Contemporary vocal performance is rarely without amplification and the employment of a microphone. The microphone has become a performance tool of the contemporary vocalist and a means for extending the voice as an instrument.

Given that the microphone is such a prevalent tool and that it is 'played' by the performer, it is possible to extend this idea to include its use as an interface for real time electronic music performance by capturing gestures via the microphone and stand, in order to derive control signals which are sent to a sound engine located in software on a computer. In the process of conceiving an alternate controller system for vocalists, it would seem attractive to address the principal limitation of the microphone/PA system as an instrument. In amplified performance situations, the vocalist has little direct control over the sound of their voice through the sound reinforcement system. Once the sound enters the microphone, any additional signal processing, such as filtering, reverberation, distortion, granulation, delay effects etc., added to the vocal signal, are usually carried out by a sound mixer or third party. Often these effects are of an intrinsically musical nature and are closely allied with other vocal production techniques employed by the performer. Wanderley and Depalle [34] showed how acoustical effects can be achieved through a performer's movements, however, the ability of the performer to shape or interact with the sound can be somewhat inhibited by the fact that some of the more critical aspects of the resulting sound are mediated by the third party.

In any proposal for an alternate controller, one must address the perceived limitations and difficulties of existing electronic music performance practice. This is particularly relevant to the recent trend towards using the laptop computer as a musical instrument without some kind of visible (to the audience) performance/ gestural interface [5], raising issues concerning the performer's relationship to the audience. The most commonly cited 'deficiency' in laptop performance is that, with the performer seated behind the laptop, there is an inherent lack of gestural communication between performer and audience due to the fact that gesture is so small and often hidden from view. As a result, the performance can have a detached, non-communicative quality.

Musical Performance is also a social act, and, whether real or virtual (in the recording studio), an audience is critical in shaping the performance event. [10]

Another problem that arises for vocalists using a laptop in performance is that the performer may be physically inhibited by the posture of sitting at the computer when trying to vocalize, see Figures 1 and 2. These issues suggest that there is a need to extend the vocalist's control over sound and to address some of the limitations of recent laptop performance practice by developing an alternate controller which not only captures gesture, but provides a visually engaging performance interface for the vocalist who wishes to work with electronic resources. The photographs in Figure 1 and 2 show both a laptop and desktop computer performance respectively, where the performer is using live vocal input. The performer is seated in front of a laptop and the microphone is placed between the performer and the computer screen.

Figure 1. Donna Hewitt Impermanent Audio 2002 [18]

Figure 2. Donna Hewitt Waveform 2001 [36]

In determining the design aspects of a proposed alternate controller, there is a need to;

a) study the gestural qualities of vocalists to identify common aspects to the 'language'

b) identify the most effective means of capturing these gestures making use of available sensing technologies and hardware

c) come to an understanding of the most effective means of mapping gesture onto sound in order to produce a flexible and playable instrument.

2. GESTURE

2.1 General Principles
The broad principles of gestural control have been discussed in the existing literature [14],[28], [29], [33]. This paper aims to focus on matters relating specifically to vocal performance and gestures relating to the development of a vocal interface. The classification of gestures used by vocal performers serves as a starting point for a more comprehensive,

rigorous analysis and categorization of gesture which will inform decisions made in relation to the more subtle aspects of controller design A preliminary categorization has been made by observing vocal performers' behaviour and drawing from personal performance experience. The most logical site for observing the movements and gestures of vocalists is popular music, where the overwhelming majority of microphone gesture practice is located. Although the gestural principles will be derived from a study of popular music, it is envisaged that it will be possible to use the Emic in a wide range of musical styles. The Emic may, in time, allow a whole new set of gestural practices to be developed, as has been seen with developments of other electronically extended musical instruments. The sensor bow described in [2] as an example, required string players to "modify their traditional technique" and showed how certain techniques were effective for the new sensor bow but not useful for playing the traditional string instrument.

Vocal performers employ a wide range of 'gestures' during performance. The term 'gesture' refers to the bodily movements allowing the performer to 'interact with their environment, to modify it and to communicate' [4].

For a vocalist, the body itself is the instrument. Playing the instrument requires control of various body parts involved with the breathing apparatus, vocal articulators and resonating cavities. Vocal performers, as with most traditional instrumentalists, also move in other ways, which may not be directly involved in sound production. "These gestures have been labeled as expressive, accompanist, ancillary or non-obvious" [34]. Studies show a wide range of expressive information is present in, and can be drawn from, the bodily gestures of a performer [7],[8],[9], [10]. Gestures provide interpretive cues for audiences and are the "by-product of psycho-physical, social and cultural practices surrounding performance". [10]

Our observations focused on how the performer approached and/or touched the microphone and microphone stand, i.e. what human movements and physical interactions with the microphone and microphone-stand commonly occurred during performance. A comprehensive understanding of the function of these 'microphone gestures' is a vast area for further study, one that will provide valuable insight into creating an effective mapping strategy.

2.2 Microphone Gesture

While every performer possesses a certain number of idiosyncratic movements and gestures, there does appear to be a number of common interactions and gestures associated with microphone and microphone stand use, which are broadly outlined below.

2.2.1 Physical Interactions

These gestures include physical interactions where the micro-phone stand is physically touched in some way. The main categories of physical interactions are the following.

2.2.1.1 Grasping Gestures

Grasping the microphone

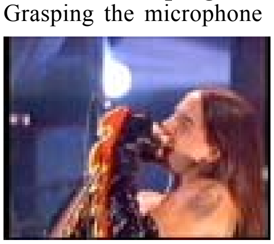

Figure 3. Red Hot Chilli Peppers, Anthony Keidis [26]

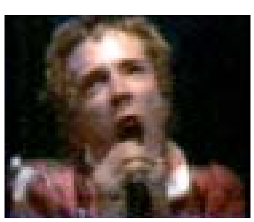

Figure 4. Sex Pistols, Johnny Rotten [1]

Grasping the stand

Figure 5. The Doors Jim Morrison [11]

Stroking the stand (sliding hands up and down the stand)

Figure 6. Mariah Carey [23]

2.2.1.2 Stand Moving Gestures

Tilting the stand

Figure 7. Red Hot Chilli Peppers [26]

Figure 8. Midnight Oil Peter Garret [31]

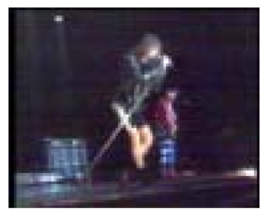

Figure 9. INXS Michael Hutchence [19]

Figure 10. The Doors Jim Morrison [11]

Moving and swinging the stand

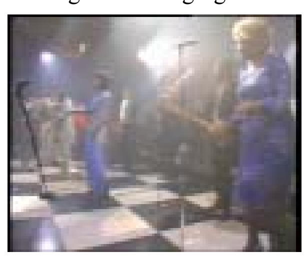

Figure 11. James Brown [30]

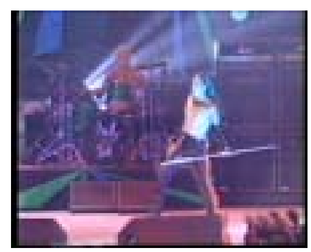

Figure 12. Red Hot Chilli Peppers [26]

(James Brown (left) throws the stand and reins it back in with the microphone lead)

Straddling the stand between the legs

Figure 13. Jim Morrison The Doors [11]

2.2.1.3 Tapping
Foot tapping the base

Hand tapping the microphone and stand

2.2.1.4 Other
Altering the stand height

Moving the microphone in and out of its clip/holder

Figure 14. Red Hot Chilli Peppers [26]

2.2.2 Free Arm/Non Contact Gestures
Vocalists make a lot of free arm gestures, where they do not touch the stand but move their arms, hands and bodies around it. In these instances, the microphone and stand provide a focal point around which the performer works or interacts, acting as a point of spatial reference for the non-contact gestures. These gestures include open hand gestures (palms facing toward the stand) and caressing type gestures (where the hands do not make contact with the stand but move around it).

Figure 15. Stevie Nicks [38]

Figure 16. Mariah Carey [23]

2.3 Functional and Contextual Aspects
Preliminary observations of vocal performers show a number of relationships between intent and physical gesture, for example, the increased grip strength of the microphone and stand most often correlate with an increase in tension in the musical intent. Performers tend to grasp the stand in a more aggressive manner when conveying more violent or angry passages, delicate stroking seems to occur more often during gentler passages. These are immediately observable one to one relationships, however gesture in performance is a complex system which is mediated by contextual factors. Gestures do not always have the same meaning or function and the mapping of each gesture needs to be considered in its unique musical context.

The congruity between gesture and intent are particularly important for a vocal performer due to the close ties between the body, psychological state and the sound produced, since “the body cues the mental representations of the music”[8]. Singers often employ learned or mimetic gestures in performance and many vocalists carefully choreograph their movements to achieve various effects such as to cue other musicians (i.e. conducting), or perhaps to create a deliberate expressive effect or to elicit a response from the audience. Singing teachers often teach singers what to do with their hands in order to furnish a performance with expressive intention. Gellrich [15] has suggested that these learned gestures can have both a positive or negative effect on a performance. It has been shown that [15] learned, mimetic or choreographed gestures can be problematic for communication with an audience when the gestures are incongruous with the intent. It may, therefore, be desirable for a performer’s gestures to appear as natural and organic as possible, however the Emic does not dictate such relationships.

While vocal performance lends itself to study, valuable insights may be gained by looking at gesture in the context of verbal communication in general. Davidson [10] has shown that singer’s gestures correspond with conversation related gestures. In gesture associated with speech, it seems that the listener relies more heavily on gesture for interpretation when the speech is ambiguous [32] or as background noise increases [27]. In relating these observations to a musical context and by noting the importance of gesture in situations with higher background noise, it is possible that in musical performance, the presence of other musical elements around the vocalist increases the importance of gesture where the intelligibility and the meaning and intent of the lyric text are of prime importance. Evidence for this can be observed in popular music performance practice, where an increase in gestural activity can be observed in situations where the vocalist is wishing to articulate the lyric clearly to the audience. Gesture is thus an important functional and expressive device in making the text intelligible for the audience and as an important means for expressive communication between the performer and audience.

3. INSTRUMENT DESIGN - Emic
Interfacing the voice and electronic processing requires a) sound capture b) a gesturing device for control input, i.e. something that the performer touches or moves to create the required control information and c) a signal processing engine which takes the captured audio and processes it in real time. The instrument is thus a gesturally responsive device bearing a relationship to an acoustic instrument in that the performer touches or moves something in order to produce and transform sound. For a vocalist, the logical signal source for an electroacoustic instrument is a microphone output. The microphone is generally placed on a microphone stand, making the microphone stand itself a logical choice as a gesturing device. The proposed design, therefore, is to build an instrument resembling a microphone stand with the parts made active as gestural controllers. The microphone stand will serve the function of a controller device, resembling a large multi-axis joystick with various buttons, sliders and sensors. The aim being to capture the common gestures that have been identified in section 2.1.

Emic is a logical adaptation of a device that many vocal performers are comfortable handling. The design aims to minimize ‘physical retraining’ by “retaining a physical interface that is functionally very similar to the practiced instrument”[14]. By choosing common gestures, the intent is to make the system more intuitive and accessible to a larger number of performers. The mic-stand is an extremely popular, widely used device that has historically endured due to its ergonomic suitability. The other advantage of the mic-stand is its familiarity to audiences. This allows the relative social codes and cultural connections associated with the interface and performance to be maintained. By re-integrating the body into the social context of music performance the device is attempting to address the perceived “lack of somatic/corporeal presence” in the performance of electro-acoustic music, an issue raised in [2].

4. MAPPING
Two fundamental approaches to the question of control mapping are those which see mapping as an integral part of the experimental process of composition and those, on the other

hand, which identify the requirement for fixed and repeatable mapping of gestural input to system control outcome. [17]

A number of points relating to mapping strategies have been highlighted in key papers and studies [17], [28], [33], [37], [39]. It seems that one of the more important aspects of mapping is to maintain the congruity between the character of the music, the expressive intent and the corresponding physical gesture. This is important from an audience perspective, since in many contexts, the audience relies on physical gesture for much of the information concerning expression and musical intent [7],[8], [9]. An important mapping consideration is to strive for a compatibility and a logical relationship between the physical gesture and the sonic outcome of that gesture and to avoid cognitive dissonances. For example, live vocalists often use the act of tilting the microphone stand when producing more intense sounds, so it would seem logical that this parameter was mapped to a parameter which intensified texture or sound intensity in some way. Stroking the microphone (ribbon sensors) would be mapped to a more intimate, subtle sound transformation. As stated in Wessell and Wright

... there should be a correspondence between the size of a control gesture and the acoustic result. Although any gesture can be mapped to any sound, instruments are most satisfying both to the performer and the audience when subtle control gestures result in subtle changes to the computers sound and larger, more forceful gestures result in more dramatic changes to the computer's sound. [37]

The two primary goals of the mapping process are firstly to have a satisfying communicative relationship from an audience perspective and secondly to create a workable relationship from a performers' perspective which meets the requirements for satisfactory control of the sound source and allows high level performance skills to be developed.

The software packages most likely to be employed in the mapping stage are Miller Puckettes' PD [25] and Ross Bencina's Audiomulch [3]. The prototype design inherently provides support for fixed approaches to control mapping while at the same time allowing new mapping strategies to be developed to support a new and emerging electro-acoustic performance practice associated with the mic-stand controller.

5. APPLICATIONS

The mic-stand interface device will find applications in a range of contemporary performance situations. Popular commercial music increasingly employs specialized vocal processing systems. This interface allows these systems to come under direct control of the performer providing scope for new avenues of musical expression.

In the field of experimental electro-acoustic music and performance art, advanced control systems have a long history. This device fits neatly into this field, where the performer is often inhibited by clumsy general purpose computing interfaces.

It may also be possible to utilize the interface in conjunction with other existing gesture capture devices such as the Yamaha MIBURI system (body suit) [40] or the Mouth Synthesizer [22] (captures facial gestures). This would enable additional gestural information to be collected.

6. PROTOTYPE FEATURES

6.1 Overview

The mic-stand interface device must provide a range of simple mechanisms to capture the characteristic gestures listed above. The control systems must be simple and intuitive but must not restrict the virtuosic performer. There is no fixed relation between control signal and sound processing.

Figure 17. Emic Prototype

6.2 Transducers

The transducers employed in the prototype are detailed below:

6.2.1 Mic Holder Joystick

The standard microphone holder allows the microphone to pivot front to back enabling the capture of microphone tilt movements. This arrangement is augmented with a dual axis pivot arrangement with a simple linear relationship between microphone angle and control signal across two orthogonal axes. The microphone holder joystick is fitted with a return spring stiff enough to support the microphone even when the stand is tilted.

Figure 18. Mic Holder Joystick

6.2.2 Slide Sensors

Microphone stand grasping and stroking gestures are captured with two 300mm linear resistive pressure/position sensors fitted either side of the stand. These sensors may be used as continuous controllers or as multi-position discrete switch inputs to be decoded in software.

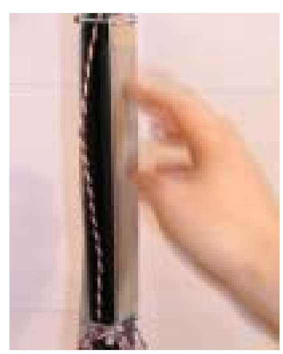

Figure 19. Right Slide Sensor

6.2.3 Distance Sensors

Free arm gestures can be captured with a distance sensor that can be played in a *Theremin* type manner. Two optical sensors with a range of 400mm are fitted just below the microphone holder on either side of the stand.

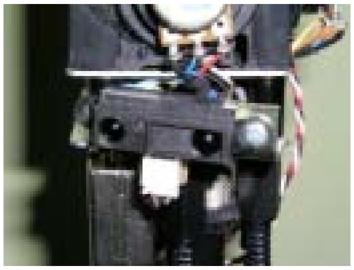

Figure 20. Distance Sensors

6.2.4 Tilt Sensors

Microphone stand tilting, swinging and moving gestures can be captured with a dual axis tilt sensor. This custom made device captures the fixed gravitational acceleration across two orthogonal axes, providing tilt sensing in X and Y planes. In addition to capturing the tilt of the stand this device will also capture rapid acceleration due to impacts on the stand from hitting, kicking or dropping.

Figure 21. Tilt Sensor

6.2.5 Mic Holder Pressure Sensors

Microphone grasping gestures can be captured with two small pressure sensors attached to the microphone holder.

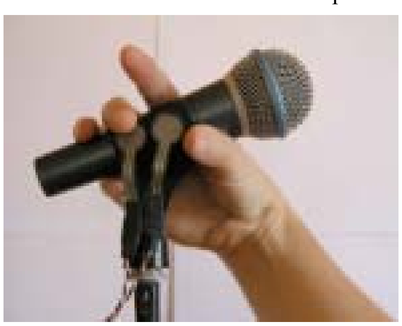

Figure 22. Mic Pressure Sensors

6.2.6 Foot Pressure Sensors

Foot pressure on the base of the stand is captured using a simple pressure sensor.

Figure 23. Foot Pressure Sensor

6.3 Control Systems

Many commercially available and experimental real-time signal-processing devices are fitted with the Musical Instrument Digital Interface (MIDI). In addition to this there are a range of commercially available analogue to MIDI interfaces. The availability and wide use of the MIDI interface is its main benefit. The most significant disadvantage of MIDI is its limited resolution. MIDI may easily be substituted with a floating-point control system such as Open Sound Control or other system specific messaging system such as Max/PD [25]. Part of the composition process will be concerned with finding mappings from available physical controls to signal processing parameters. The system must be flexible in respect of providing unlimited mapping arrangements.

6.4 Interfacing

The prototype system employs a simple interfacing strategy based on ready availability of components, simplicity, tourability and reliability in performance environments. The components can be easily removed to protect them while travelling.

6.4.1 Multi-core Cable

This simple interfacing method provides reliability and low cost construction. FM radio data transmission devices may replace this method.

6.4.2 CV to MIDI Converter

A low cost control voltage to MIDI converter made by Angelo Fraietta [13] is used. This device provides sixteen inputs for analog to MIDI conversion.

6.5 Control Feedback

The tactile controls employed in the system provide inherent positional feedback. The choice of non-mechanical sensing technology in the tilt sensor provides the user with the familiar ballistic response associated with conventional microphone stands. The non-tactile distance sensor provides no positional feedback. These sensors require either advanced technique or reduction of sensitivity or resolution in the control mapping stage. Slide position sensors are fitted with tactile detents for positional orientation or multi-position switch use.

7. CONCLUSION

Having identified the key categories of gesture and the means by which those gestures may be captured, an Emic prototype has been developed. The next stage in the process is to develop workable mapping strategies and to implement the compositional process.

8. REFERENCES

[1] Anarchy in the U.K.. The Filth and The Fury – A Sex Pistols Film. Video. Directed by Julien Temple. USA, Warner Home Video (2000).

[2] Bahn, C., Hahn,T. & Trueman, D. Physicality and Feedback: A Focus on the Body in the Performance of Electronic Music. Multimedia Technology and Applications Conference. University of California, Irvine. (2001)

[3] Bencina, R. Audiomulch: Interactive Music Studio (software) http://www.audiomulch.com

[4] Cadoz, C., Luciani, A., & Florens, J. L. Responsive input devices and sound synthesis by simulation of instrumental mechanisms: The CORDIS system. Computer Music Journal, 8(3) (1984), 60-73.

[5] Cascone, K., Laptop Music: Counterfeiting Aura in the Age of Reproduction. Parachute No. 107 (2002), 56.

[6] Cook, P., Principles for Designing Computer Music Controllers. Conference Proceedings NIME (2001) http://www.csl.sony.co.jp/person/poup/research/chi2000wshp/papers/cook.pdf

[7] Davidson, J.W. Visual perception of performance manner in the movements of solo musicians. Psychology of Music*,* 21 (1993), 103-13.

[8] Davidson, J.W. Which area of a pianist's body convey information about expressive intention to an audience? Journal of Human Movement Studies, 26 (1994), 279-301.

[9] Davidson, J.W. What does the visual information contained in music performances offer the observer? Some preliminary thoughts. R. Steinberg (ed.), The Music Machine: Psychophysiology and Psychopathology of the Sense of Music Heidelberg, Germany: Springer (1995), 103-15.

[10] Davidson, J. W. The role of the body in the production and perception of solo vocal performance: A case study of Annie Lennox. Musicae Scientiae, Vol V, No. 2, (Fall 2001), 235-256.

[11] The Doors, Live at the Hollywood Bowl. Video. Directed by Ray Manzarek USA. Universal Edition (1987).

[12] Ekman, P. & Friesen, W. The repertoire of nonverbal behavioral categories – origins, usage, and coding. Semiotica, 1 (1969), 49-98.

[13] Fraietta, A. CV-to Midi–Midi to CV Midi Controller http://www.users.bigpond.com/angelo_f/midicontroller/MidiController.html

[14] Garnett, G., & Goudeseune, C. Performance Factors in Control of High-Dimensional Spaces. Proceedings of the 1999 International Computer Music Conference. San Francisco. (1999), 268–271 http://zx81.isl.uiuc.edu/camilleg/icmc99.html

[15] Gellrich, M. Concentration and Tension. British Journal of Music Education, 8 (1991), 167-79.

[16] Hunt, A., & Kirk, R. Mapping Strategies for Musical Performance. M. Wanderley and M. Battier (eds). Trends in Gestural Control of Music. Ircam, 2000.

[17] Hunt, A., Wanderley, M. & Kirk, R. Towards a Model for Instrumental mapping in Expert Musical Interaction. In the Proceedings of the International Computer Music Conference. San Fransisco. International Computer Music Association (2000), 209-212.

[18] Impermanent Audio, Live Performance. Frequency Lab. Sydney, Australia. (August 2002). Photo : Mr Snow. http://laudible.net/impaud/archive/

[19] INXS : In Search of Excellence (1989) Video Directed by Hamish Cameron. Polygram Hamrod Productions ISBN 080 850 *

[20] Jorda, S. FMOL: Toward User-Friendly, Sophisticated New Musical Instruments. Computer Music Journal 25:3 (Fall 2002), 23-29.

[21] Lazzetta, F. Meaning in Musical Gesture. In Trends in Gestural Control of Music. M. Wanderley and M. Battier (eds) Ircam - Centre Pompidou, 2000.

[22] Lyons, M. & Tetstani, N. Mouthesisor A Facial Gesture Musical Interface http://www.mis.atr.co.jp/~mlyons/pub_pdf/79.pdf

[23] Mariah Carey performing Hero, Live at Proctors Theater New York City avail on NBC Thanksgiving Special Home Video Video. Directed by Larry Jordan USA 1994 Warner Brothers.

[24] Mulder, A., Fels, S. & Mase, K. Mapping Virtual Object Manipulation to Sound Variation. T. Rai and R. Basset (eds). IPSJ SIG notes 97(122), (1997). 63-68.

[25] Puckette, M. Pure Data (software) http://www.crca.ucsd.edu/~msp/software.html

[26] Red Hot Chilli Peppers 'Warped' (live performance) MTV Video Music Awards (1995) http://www.mtv.com/bands/az/red_hot_chili_peppers/audvid.jhtml

[27] Rodgers, W.T. The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances. Human Communication Research, 5 (1978), 54-62

[28] Rovan, J., Wanderley, M., Dubnov, S., & Depalle, P. Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance. Kansei, The Technology of Emotion. Proceedings of the AIMI International Workshop, A. Camurri, (ed.) Genoa: Associazione di Informatica Musicale Italiana, October 3-4, (1997), 68-73. http://www.ircam.fr/equipes/analyse-synthese/wanderle/Gestes/Externe/Mapp/kansei_final.html

[29] Sapir, S. Gestural Control of Digital Audio Environments. Journal of New Music Research, Vol. 31, No. 2 (2002), 119-129.

[30] A Soul Session, James Brown and Friends. Video. Directed by David Grossman, USA, Cinemax, 1987.

[31] So You Want To Be a Rock Star. Music Documentary Video. Directed by Ray Argall, Sydney Australia. Produced by Christina Pozzan, Daniel Scharf, Ray Argall & Australian Film Commission, (1987).

[32] Thompson, L.A. & Massaro, D.W. Evaluation and integration of speech and pointing gestures during referential understanding. Journal of Experimental Child Psychology, 42 (1986), 144-168.

[33] Wanderley, M. Gestural Control of Music International Workshop Human Supervision and Control in Engineering and Music**,** Kassel, Germany (September 2001).

[34] Wanderley, M & Depalle, P. Gesturally-Controlled Digital Audio Effects. In Proceedings of the Cost G-6 Conference on digital Audio Effects (DAFX-01), Limerick, Ireland, (Dec-2001).

[35] Wanderley, M. Quantitative Analysis of Non-obvious Performer Gestures. Gesture and Sign Language in Human-Computer Interaction. Wachsmuth, I & Sowa, T. (eds.). Springer Verlag, 2002.

[36] 'Waveform' Concert. Australian Computer Music Conference. University of Western Sydney, Australia (July, 2001). http://www.acma.asn.au/

[37] Wessel, D., & Wright, M. Problems and Prospects for Intimate Musical Control of Computers. Computer Music Journal, 26:3 (Fall 2002), 11-22. http://www.csl.sony.co.jp/person/poup/research/chi2000wshp/papers/wessel.pdf

[38] White Wing Dove. Video. Directed by Marty Callner CBS/FOX, U.K. (1983).

[39] Winkler, T. Making Motion Musical: Gestural Mapping Strategies for Interactive Computer Music. Proceedings of the 1995 International Computer Music Conference. San Francisco, International Computer Music Association (1995), 261-264.

[40] Yamaha, (1996) Miburi R3 Manual. Tokyo, Japan, Yamaha Corporation (In Japanese, no author listed)