USING MOTION CAPTURE DATA IN PERCEPTION STUDIES

Jonna K. Vuoskoski

EMBODIED APPROACH TO MUSIC COGNITION

- Body movement is integral to the production and perception of music
  - Serves many functions besides sound production; communication, expression, cognition…
- Making and perceiving music are auditory–visual–kinaesthetic processes
  - Music perception is multimodal; we perceive music with the help of visual/kinematic information and effort/dynamics sensations (Godøy, 2003)

EMBODIED COGNITION

- Cognition is shaped by aspects of the body and the environment (e.g., Shapiro, 2007)
  - Concepts and categories, cognitive and perceptual processes…
- Cognition primarily serves action
- Emphasizes the embodied and situated nature of cognition
- Dynamic interaction between minds, bodies, and their environment

EMBODIED MUSIC COGNITION

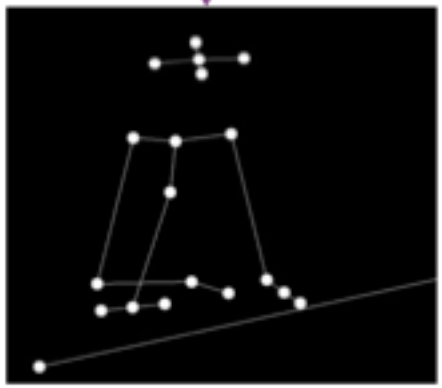
- Musical sound not only perceived as an auditory signal, but also as intentional, expressive motor acts behind the signal (Molnar-Szakacs &Overy, 2006)
  - Perceived physical energy in sound is converted into 'action-oriented' meanings through correspondence with bodily movement and gesture (Leman, 2008)
  - Music utilises similar acoustic cues as human vocal expression of emotions (Juslin & Laukka, 2003), and emulates the speed, trajectory and smoothness/jerkiness of human movement and gestures (Jackendoff & Lerdahl, 2006)
- Perception and action are coupled; we understand musical sounds and actions through motor simulation

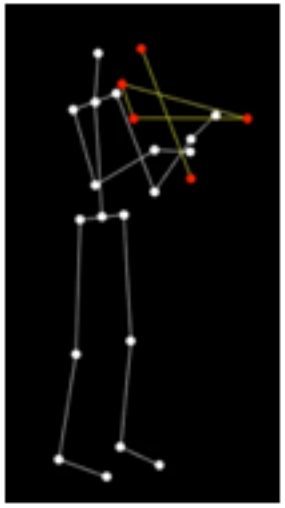
MUSIC – A MULTISENSORY PHENOMENON

- The body movements and gestures of the performer constitute an important source of information for audiences
  - Different expressive intentions are better recognised based on visual rather than auditory information (Davidson, 1993, 1994)
  - Non-experts can more reliably guess the winners of classical music competitions based on visual rather than auditory (or audiovisual!) information (Tsay, 2013)
- We have a natural, automatic, and nonconscious dependence on visual cues

WHAT CAN MOTION CAPTURE OFFER TO PERCEPTUAL STUDIES?

- Abstraction of body movement
  - Enhanced focus on movement
  - Controlling for variability in appearance
- More options/opportunities for manipulation
  - Multiple options for animations/visualisations
  - Time-warping
  - Movement synthesis
  - Aggregated/averaged movement





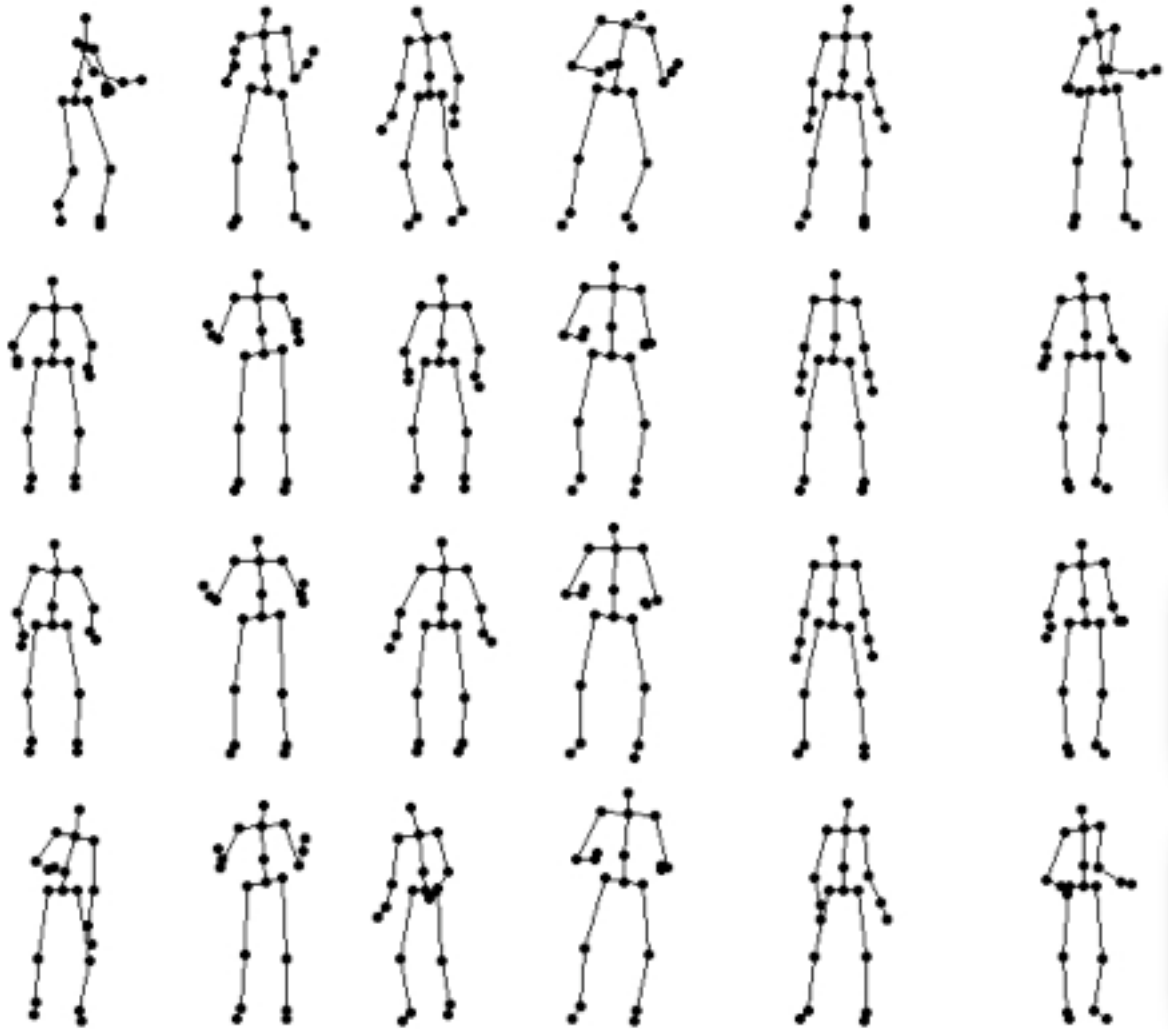
EXAMPLE: DECOMPOSITION OF MOVEMENT PERIODICITY BY FILTERING

Original

f = 2 Hz

f = 1 Hz

f = 0.5 Hz



vertical movement

limb swing

body sway

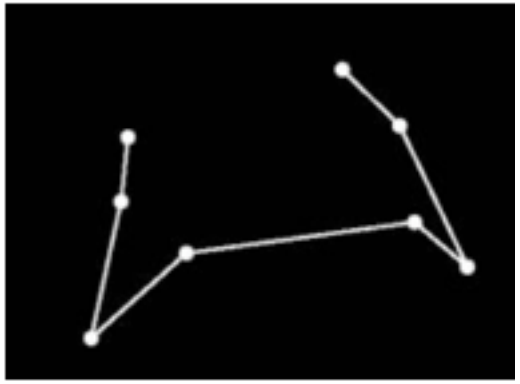
Toiviainen, P., Luck, G. & Thompson, M. (2010). Embodied meter: hierarchical eigenmodes in music-induced movement. Music Perception, 28(1), 59-70.

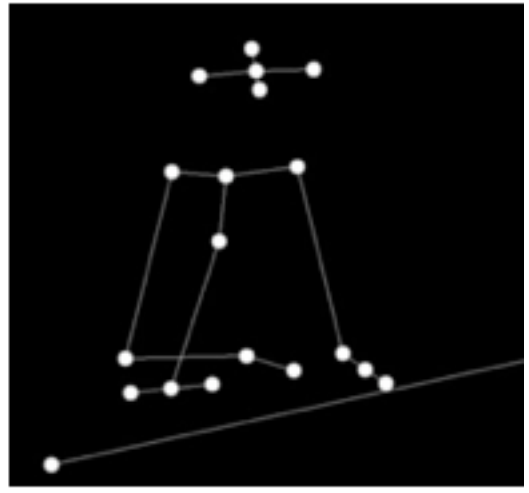
WHAT CAN MOTION CAPTURE OFFER TO PERCEPTUAL STUDIES (2)

- Associating particular movement features to perceptual phenomena/attributes
  - Predicting observer ratings using a set of movement features derived from mocap data
  - Manipulating certain movement features in mocap animations and measuring their effect on observer ratings
- Exploring the salience of visual kinematic information, investigating the boundary conditions for accurate decoding
- Exploring cross-modal interactions
  - Manipulating visual and auditory components independently

MUSIC PERCEPTION STUDIES USING MOTION CAPTURE ANIMATIONS

- Perceived expressivity in conductor's gestures (Luck, Toiviainen, & Thompson, 2010)
- Perceived emotion in non-expert dancers' movements (Burger et al., 2013)
- Crossmodal effects of dancers' movements on perceived tempo of music (London et al., 2016)
- Crossmodal (audiovisual) interactions in the perception of piano performance (Vuoskoski et al., 2013; 2016)





EXAMPLE: EXPLORING AUDIOVISUAL INTERACTIONS IN MUSICAL PERFORMANCE

MULTISENSORY INTEGRATION

- Describes how humans form coherent perceptions by combining sensory information from various modalities, and how different sensory modalities interact
- Sensory illusions; the McGurk effect (McGurk & MacDonald, 1976)
- Optimal sensory integration: more weight is given to the modality that provides the more reliable sensory information (Alais & Burr, 2004; Ernst & Banks, 2002)
- Unity assumption; degree of sensory integration depends on the detection of a causal link between the stimuli (de Gelder, Pourtois, & Weiskrantz, 2002; Schutz & Kubovy, 2009)

THE MCGURK EFFECT

- The motor theory of speech perception:
  - People perceive spoken words by identifying the vocal tract gestures with which they are pronounced rather than by identifying the sound patterns (e.g., Liberman et al., 1967)

CROSSMODAL INTERACTIONS IN MUSIC

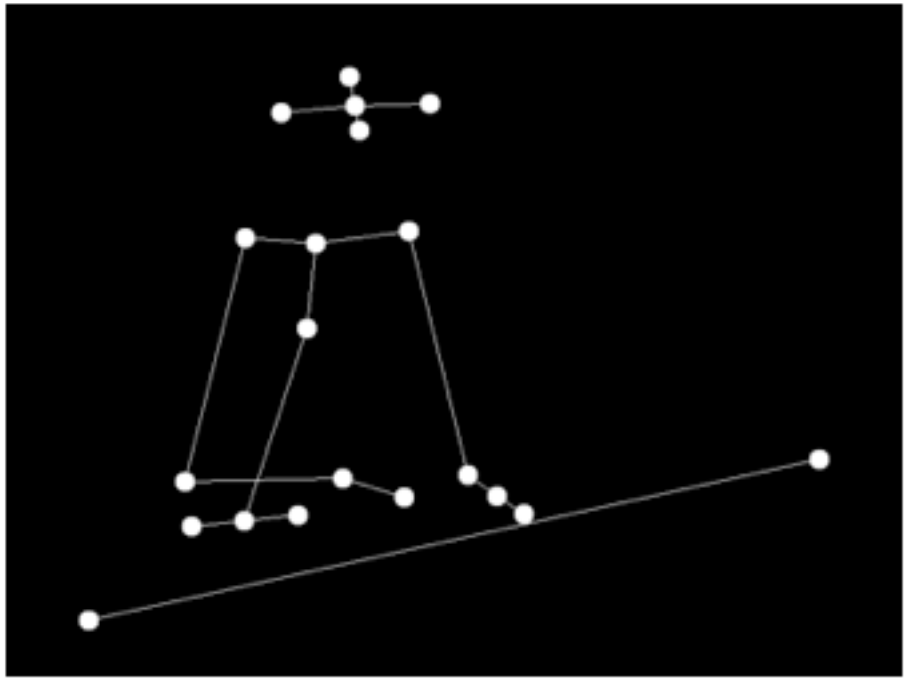
- Audiovisual interactions have been demonstrated in the perception of:
  - Interval size & affect (Quinto et al., 2010; Thompson et al., 2005)
  - Note timbre & duration (Saldaña & Rosenblum, 1993; Schutz & Lipscomb, 2007; Schutz & Kubovy, 2009)
- Method: both modalities are manipulated independently, resulting resulting in matching and mismatching bimodal pairings of stimuli
  - E.g., short & long marimba sounds paired with *both* short & long striking gestures
  - ➡ Enables the disambiguation of the effects of visual and auditory information

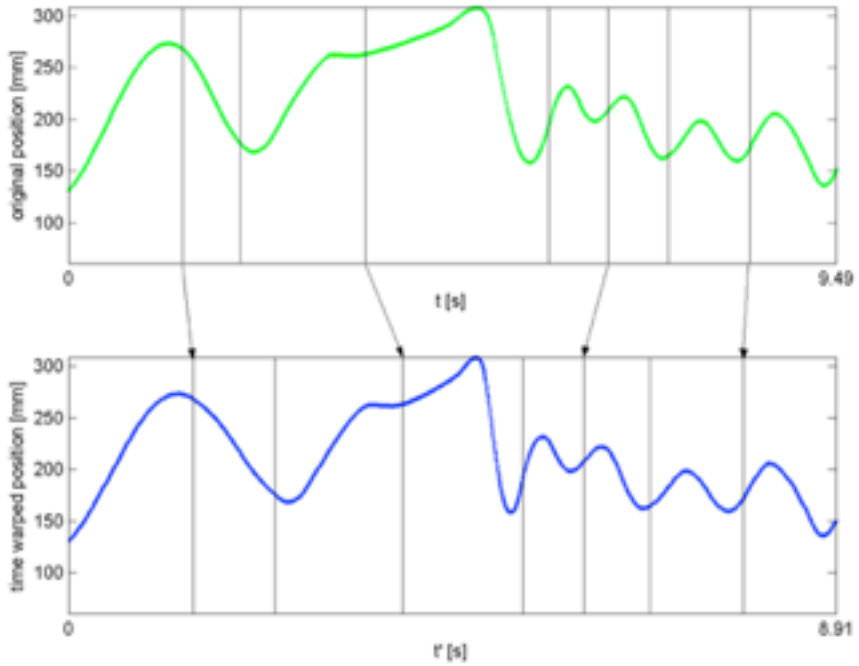
AUDIOVISUAL INTERACTIONS IN COMPLEX MUSICAL MATERIAL?

- Independent manipulation of the auditory and visual components of musical performance difficult because of variation in timing
  - Some solutions:
    - A constant auditory stimulus combined with visual information showing actors portraying different expressive intentions (e.g., Juchniewicz, 2008; Morrison et al., 2009)
    - Temporally unaligned combinations of short musical improvisations (Petrini et al., 2010)
    - Temporally unaligned combinations of happy an sad clarinet performances (with fingers obscured from view; Krahé et al., 2013)
    - Asking performers to alter movement while performing (Thompson & Luck, 2012)

SOLUTION?

Motion capture + Time-warping

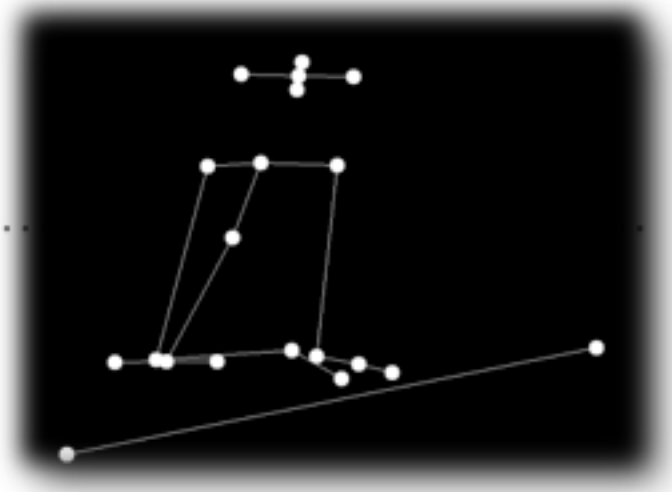




STUDY QUESTIONS

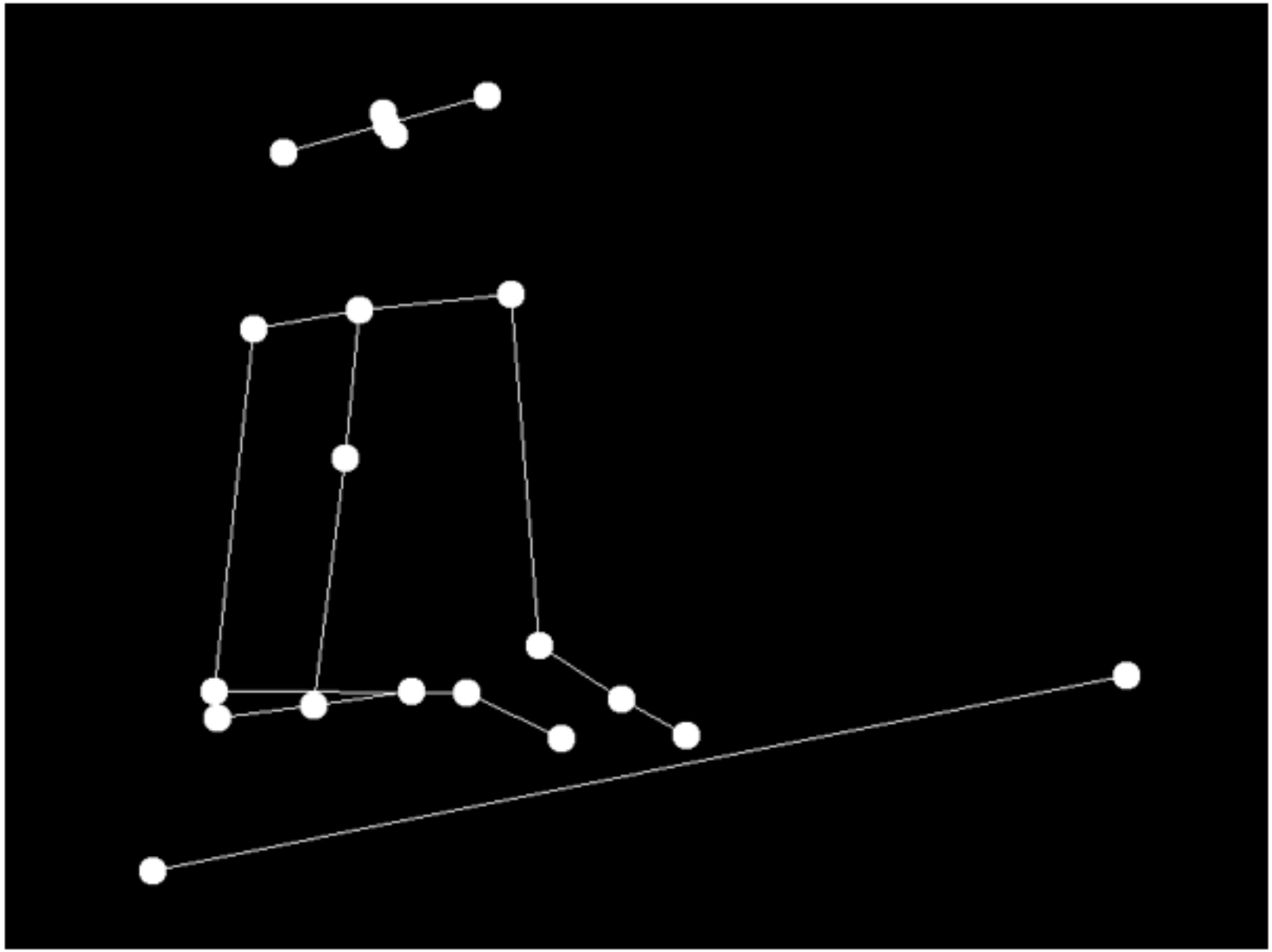
- What are the relative contributions of auditory and visual performance cues to perceived expressivity, and the emotional impact of a performance?
- Do visual movement cues have crossmodal effects on the perception of auditory musical features (e.g., loudness variability)?

STIMULUS MATERIAL



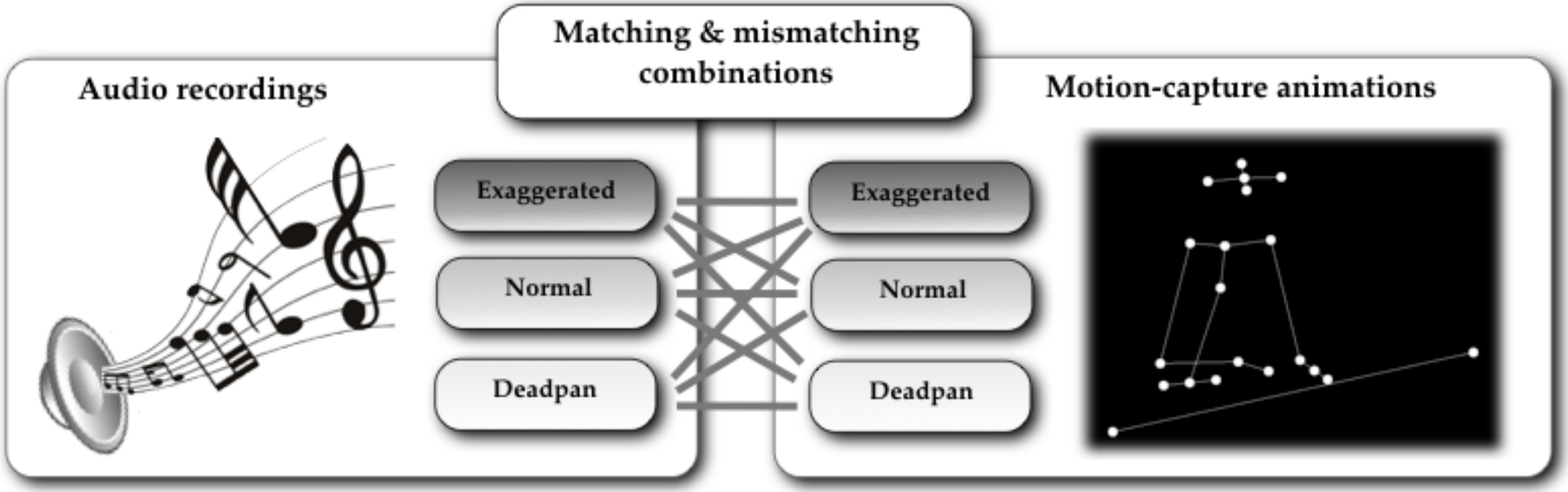
- 2 pianists (1 male and 1 female)
- Each performed Chopin's Prelude in E minor (Op. 28, No. 4) 3 times with different expressive intentions:
  - Deadpan, Normal, and Exaggerated
- Pianists' movements were recorded using an optical 8-camera motion-capture system (Qualisys; sampling rate 120 Hz)

Pianist 2 - Exaggerated performance



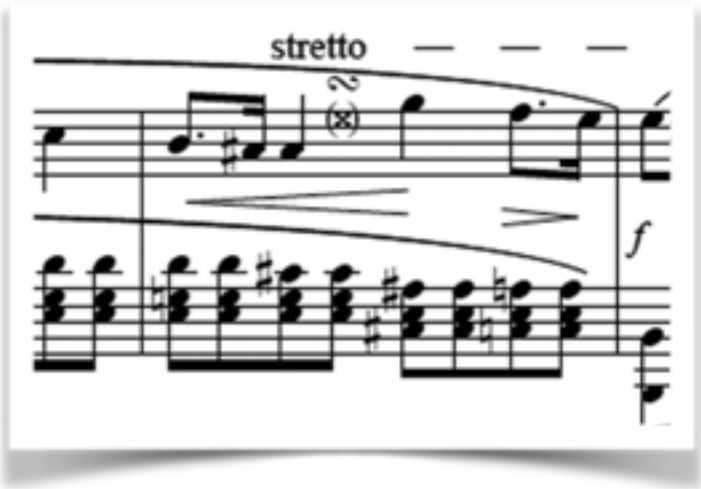
METHOD

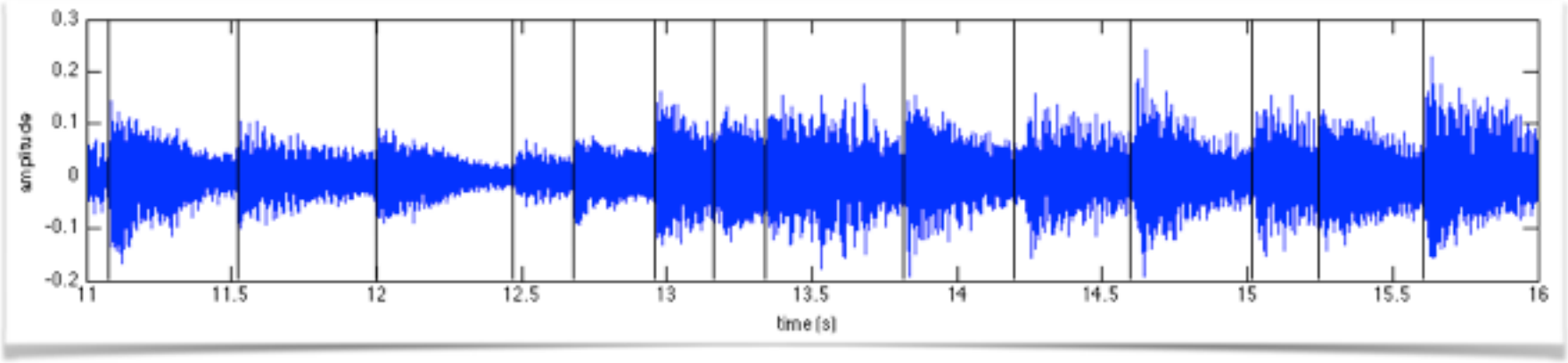
- Mismatching audio-visual combinations were generated by aligning the motion-capture data to the audio using time-warping algorithms



ANNOTATION

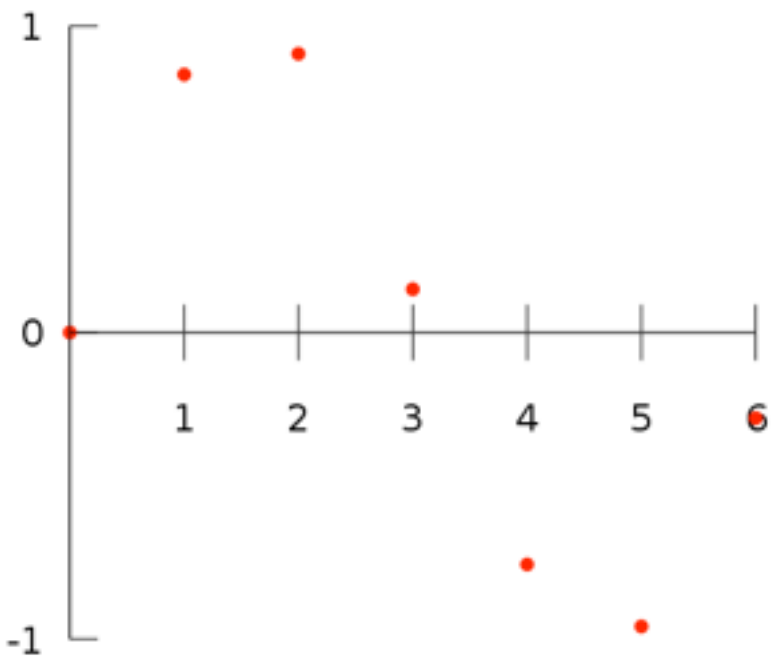
Note-onsets were annotated to generate timing profiles for each performance

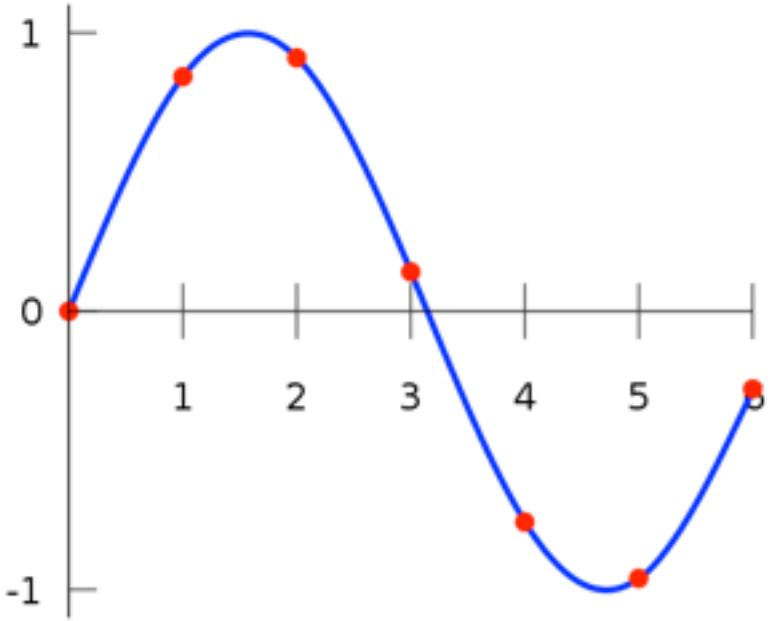




TIME-WARPING

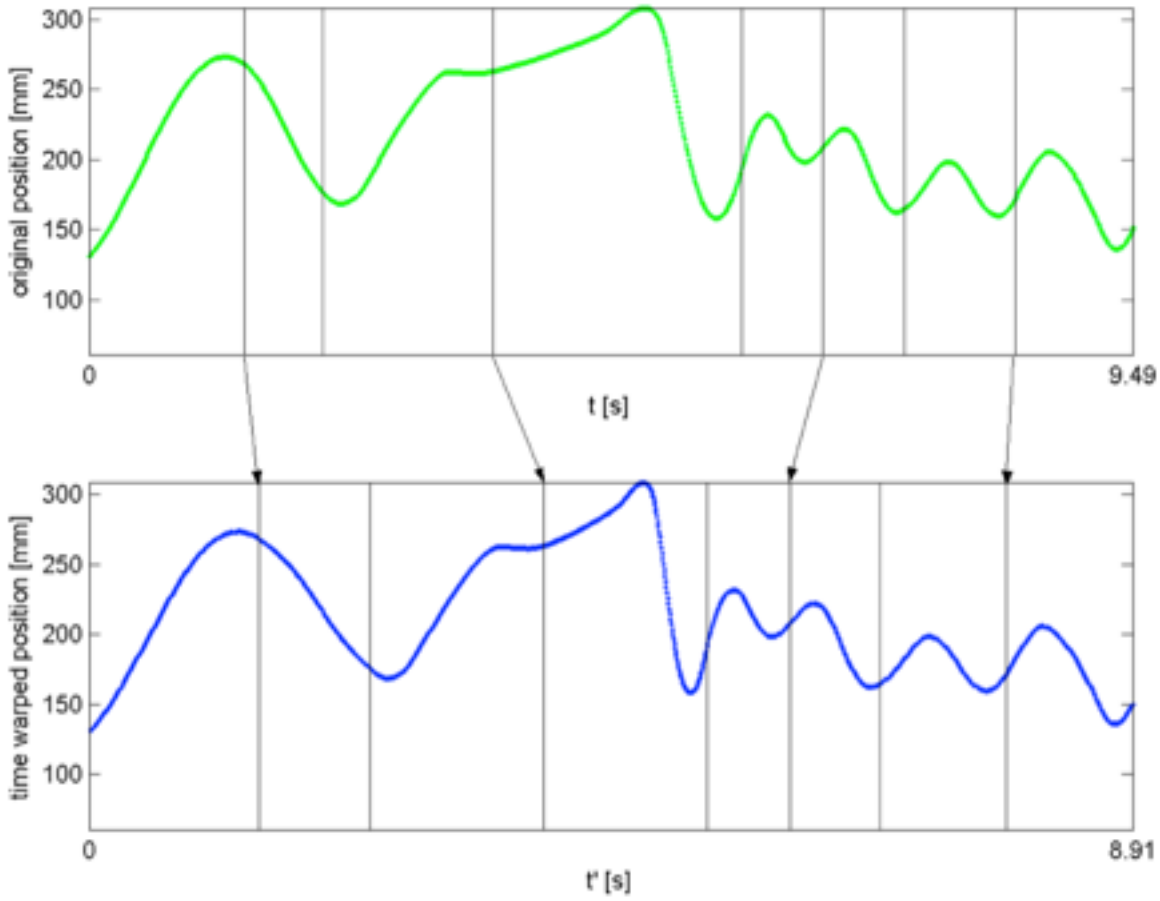
- The motion-capture data was functionalized using cubic splines (spline interpolation)
- Script based on the work of Wanderley et al. (2005) and Verron (2005) on the analysis of motion-capture data





TIME-WARPING (2)

- Using the annotated timing profiles as templates, curve-stretching algorithms were then applied to the splines between each note onset

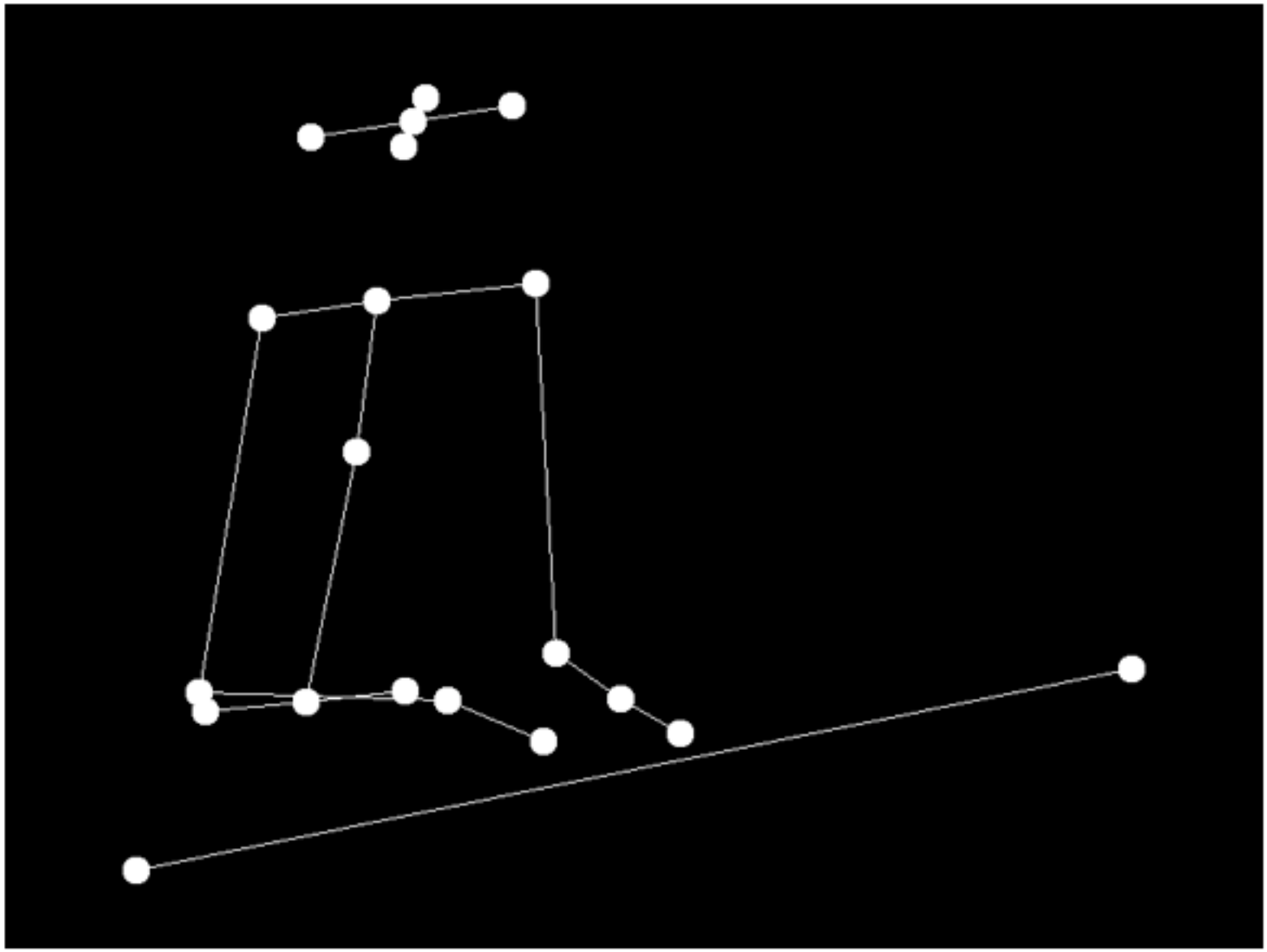


Verron, 2005

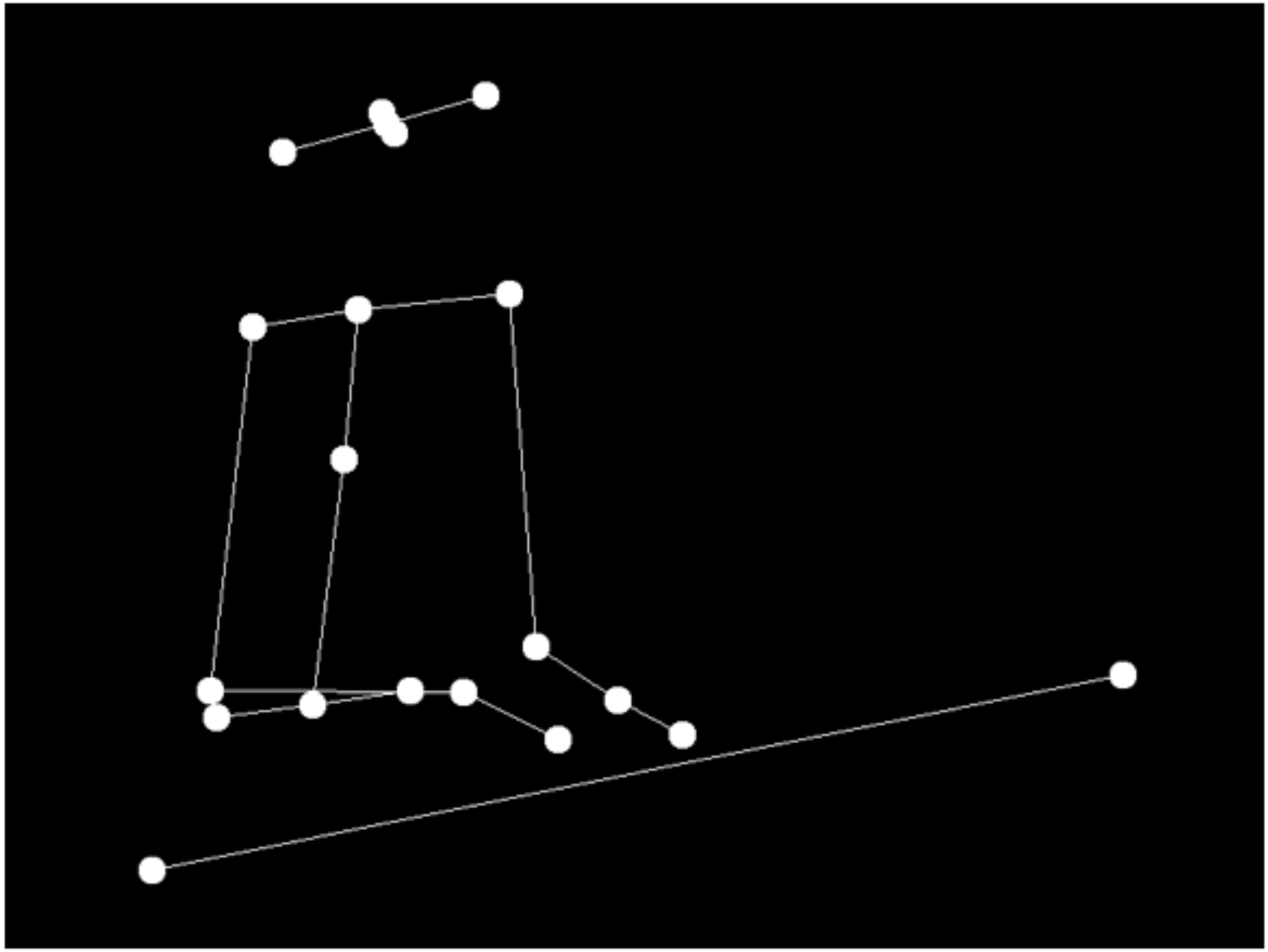
TIME-WARPING (3)

- Finally, the time-warped curves were sampled to create point-light animations
  - MoCap-toolbox for MatLab (Toiviainen & Burger, 2011)

Normal audio – deadpan movement

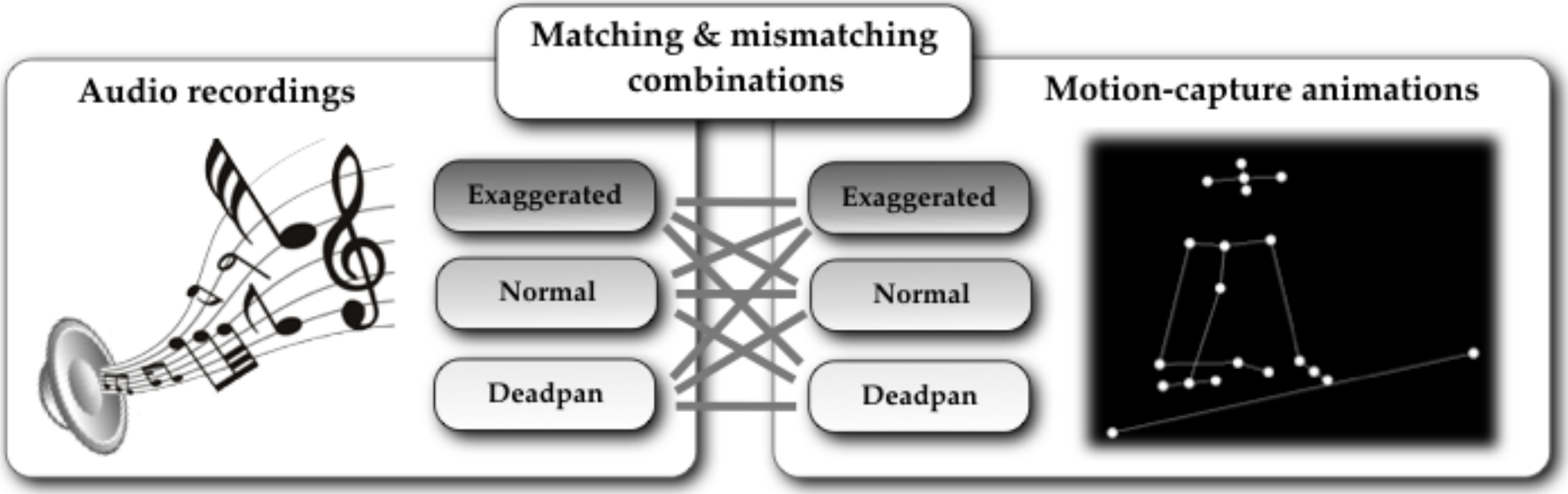


Normal audio – exaggerated movement



STIMULI

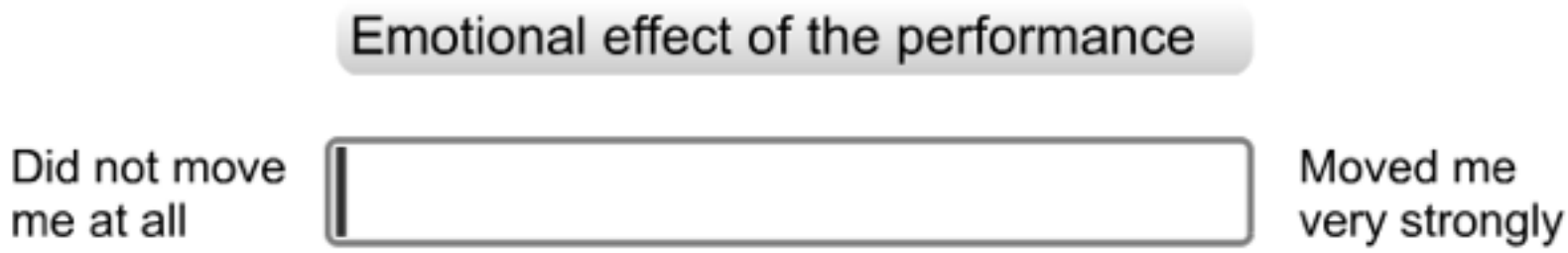
- 18 audiovisual excerpts; length approx. 30 s



  - 6 audio-only excerpts
  - 18 video-only excerpts

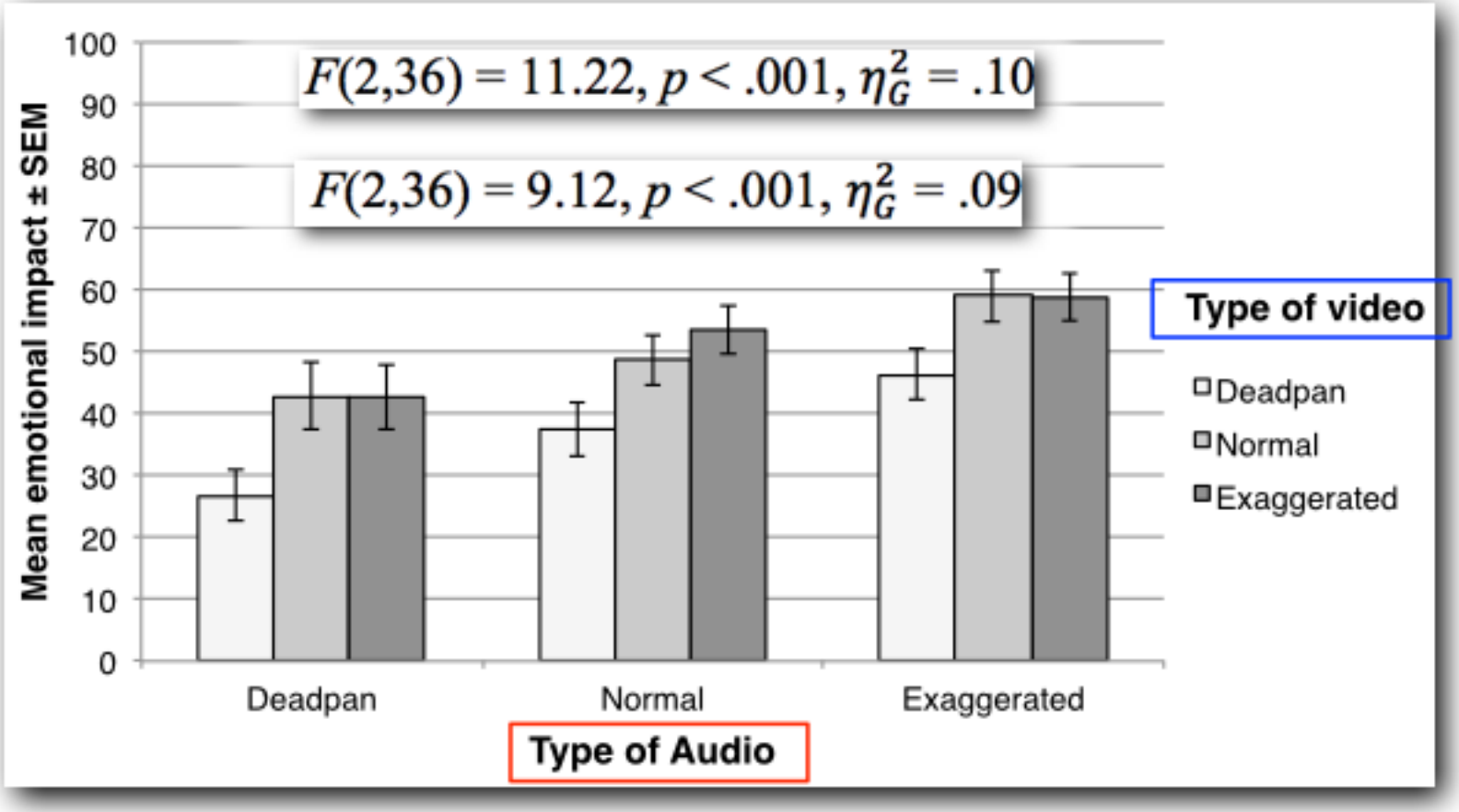
PERCEPTUAL EXPERIMENTS: EXAMPLE

- Felt emotional impact
  - N = 19
  - Aim: to investigate the relative contributions of auditory and visual cues to felt emotional impact



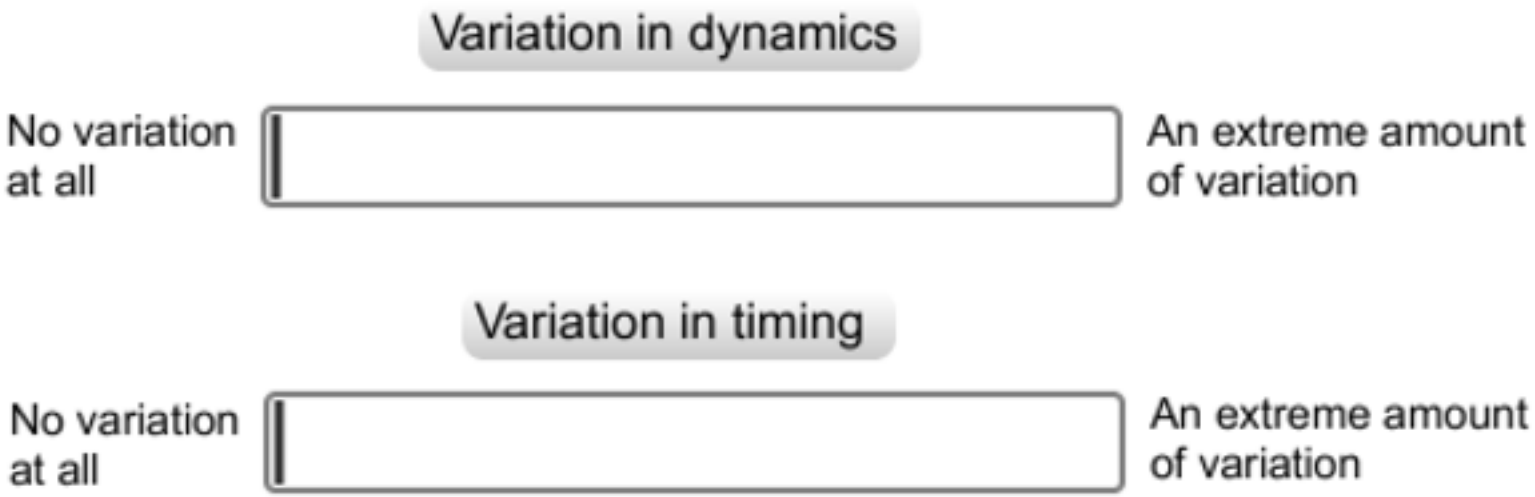
RESULTS: FELT EMOTIONAL IMPACT

- Auditory performance cues and visual movement cues both contributed significantly to participants' subjective emotional responses



METHOD

- **Experiment 2**: perception of expressive musical features
  - 17 participants, all with musical training
  - Aim: to investigate the effects of visual information on the perception of loudness and timing variability

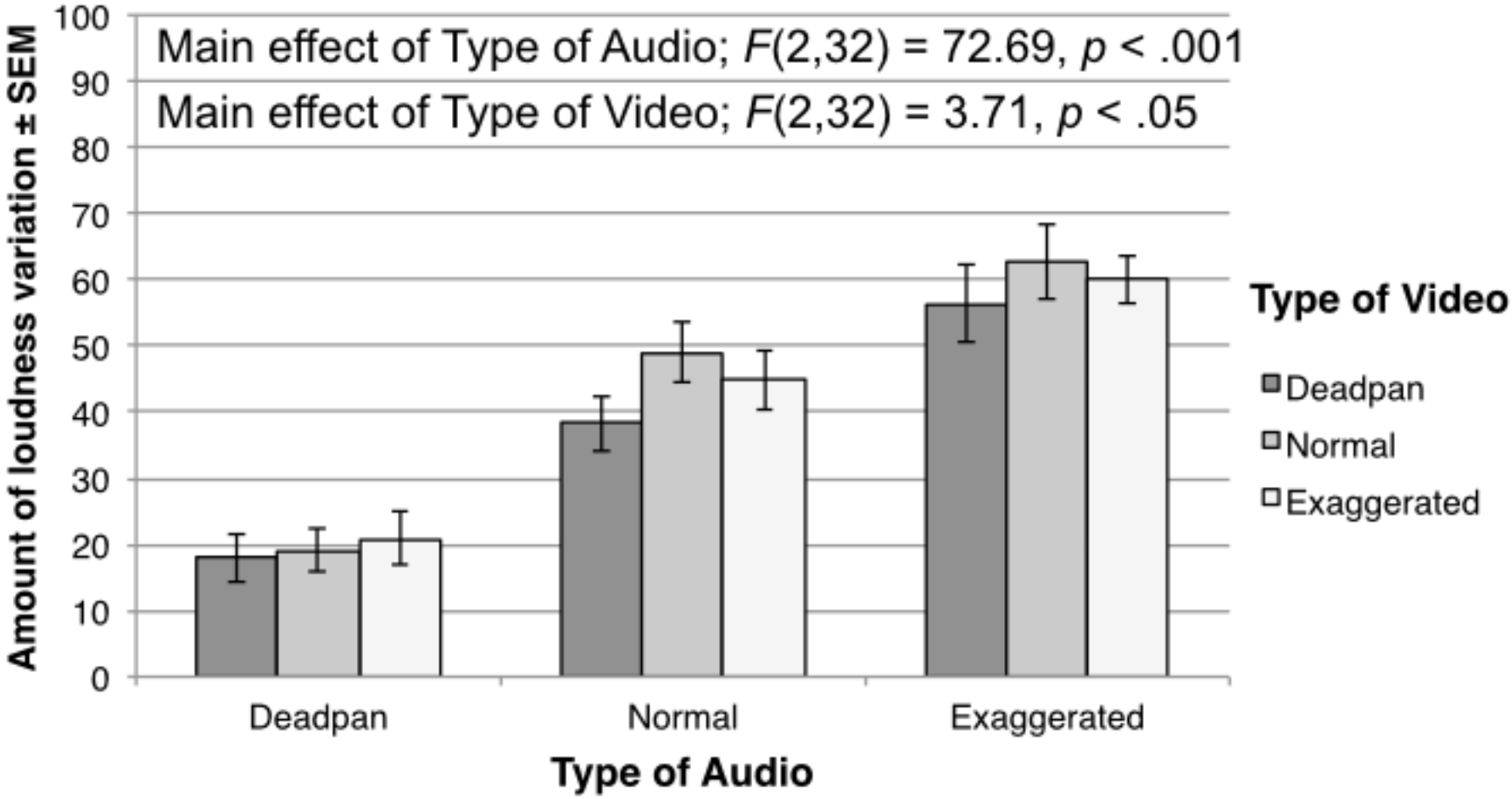


OBJECTIVE MEASURES OF TEMPO AND LOUDNESS VARIABILITY

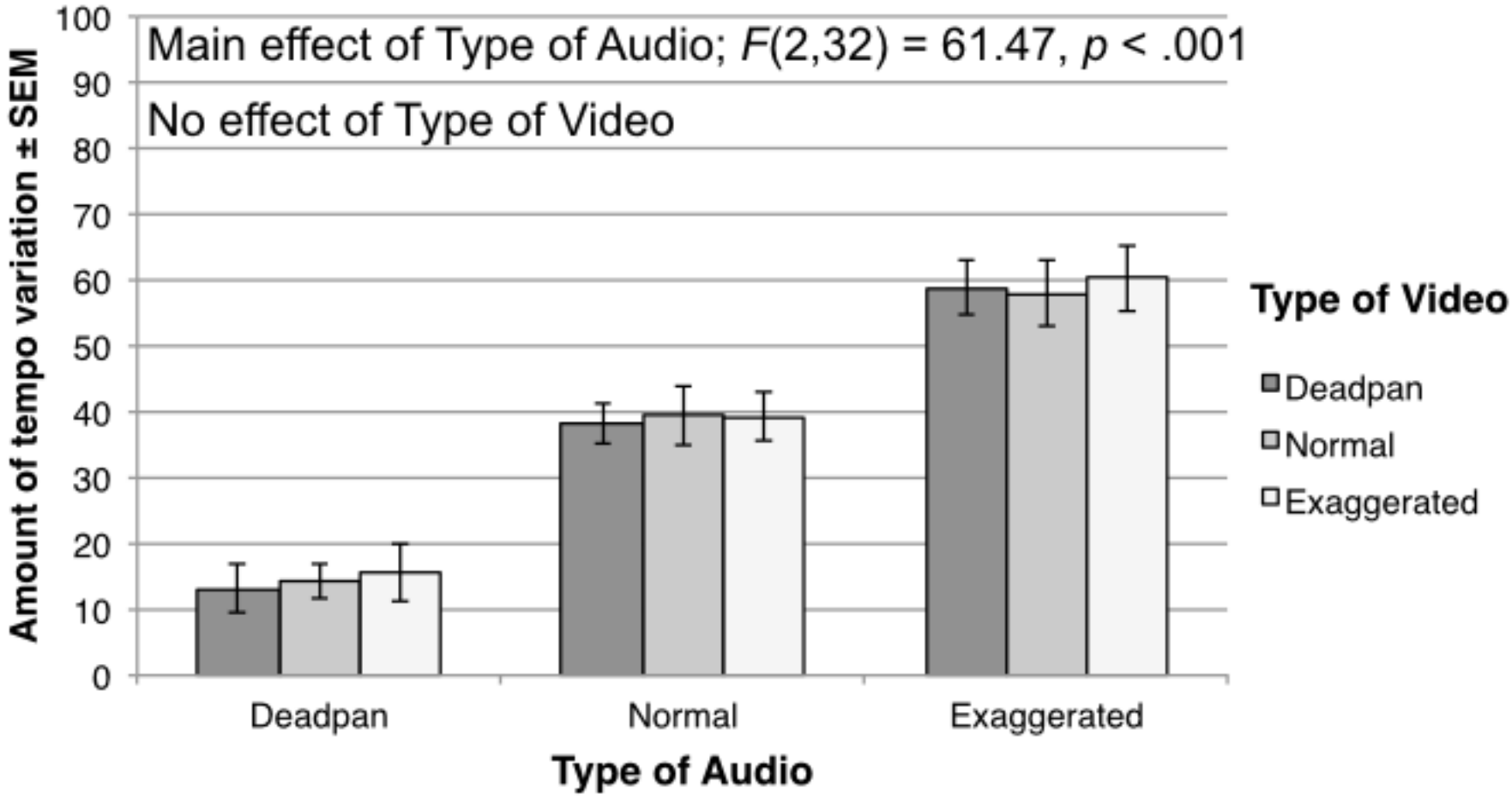
	Performance type	Mean tempo (bpm)	Tempo variability (%)	Mean RMS (SD)*	Amount of movement (m)
Pianist 1	Deadpan	65.48	5.84	2.05 (0.79)	15.69
	Normal	62.22	15.08	2.45 (1.22)	36.7
	Exaggerated	58.87	17.31	3.23 (1.50)	44.03
Pianist 2	Deadpan	59.29	8.24	1.72 (0.73)	18.27
	Normal	58.72	16.18	2.23 (1.30)	29.61
	Exaggerated	57.39	24.26	2.32 (1.54)	33.36

*RMS values and standard deviations have been multiplied by 1000

AUDIOVISUAL CONDITION: LOUDNESS VARIABILITY



AUDIOVISUAL CONDITION: TIMING VARIABILITY



CONCLUSIONS

- For emotional responses evoked by musical performances, performers' expressive body movements appear to be as important as musical expression
- Visual information had a statistically significant effect on the ratings of loudness variability (but not timing variability)
  - Evidence of crossmodal effects of the size of pianists movements on loudness perception?
  - The lack of effects on ratings of tempo variability suggests that the effects cannot be explained simply in terms of response bias

Atten Percept Psychophys
DOI 10.3758/s13414-013-0582-2

Crossmodal interactions in the perception of expressivity in musical performance

Jonna K. Vuoskoski · Marc R. Thompson · Eric F. Clarke · Charles Spence

Abstract In musical performance, bodily gestures play an important role in communicating expressive intentions to audiences. Although previous studies have demonstrated that visual information can have an effect on the perceived expressivity of musical performances, the investigation of audiovisual interactions has been held back by the technical difficulties associated with the generation of controlled, mismatching stimuli. With the present study, we aimed to address this issue by utilizing a novel method in order to generate controlled, balanced stimuli that comprised both matching and mismatching bimodal combinations of different expressive intentions. The aim of Experiment 1 was to investigate the auditory and visual expressivity. In certain performance conditions, visual cues had an effect on the ratings of auditory expressivity, and auditory cues had a small effect on the ratings of visual expressivity.

Keywords Crossmodal interaction · Multisensory integration · Music cognition · Performance · Expressivity · Gesture

Gestures constitute an integral part of human communication. They can facilitate the comprehension of speech, can convey

music perception

Vol. 33 No. 4, April 2016

Interaction of Sight and Sound in the Perception and Experience of Musical Performance

Jonna K. Vuoskoski
University of Oxford, Oxford, United Kingdom

Marc R. Thompson
University of Jyväskylä, Jyväskylä, Finland

Charles Spence, & Eric F. Clarke
University of Oxford, Oxford, United Kingdom

Recently, Vuoskoski, Thompson, Clarke, and Spence (2014) demonstrated that visual kinematic performance cues may be more important than auditory performance cues in terms of observers' ratings of expressivity perceived in audiovisual excerpts of piano playing, and that visual kinematic performance cues had crossmodal effects on the perception of auditory expressivity. The present study was designed to extend these findings, and to provide additional information about the roles of sight and sound in the perception

Music is an inherently multisensory phenomenon, comprising auditory, visual, and somatosensory components. In musical performance, a performer's body movements and gestures can convey a range of meaningful information to audiences and co-performers alike, including emotional expression (Castellano, Mortillaro, Camurri, Volpe, & Scherer, 2008; Dahl & Friberg, 2007; Davidson, 1993, 1994) and phrasing (Juchniewicz, 2008; Vines, Krumhansl, Wanderley, & Levitin, 2006), as well as musical ideas and timing (Glowinski et al., 2013; Goebl & Palmer, 2009; Williamon & Davidson, 2002). The salience of visual kinematic information (i.e., visual information about performers' body movements and gestures) for an observer's perception and experience of a musical performance has been widely documented (e.g., Chapados & Levitin, 2008; Davidson, 1993; Tsay, 2013; Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011; Vines et al., 2006), and a recent meta-analysis

THANK YOU!

+REFERENCES (1/2)

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), 257-262.

Burger, B. & Toiviainen, P. (2013). MoCap Toolbox – A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), Proceedings of the 10th Sound and Music Computing Conference, (SMC). Stockholm, Sweden: KTH Royal Institute of Technology.

Davidson, J. W. (1993). Visual perception and performance manner in the movements of solo musicians. Psychology of Music, 21, 103-113.

Davidson, J. W. (1994). What type of information is conveyed in the body movements of solo musician performers? Journal of Human Movement Science, 6, 279-301.

De Gelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of Sciences*, *99*(6), 4121-4126.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429.

Godøy, R. I. (2003). Motor-mimetic music cognition. *Leonardo*, *36*(4), 317-319.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it?. *Cognition*, *100*(1), 33-72.

Juchniewicz, J. (2008). The influence of physical movement on the perception of musical performance. *Psychology of Music,* 36(4), 417-427.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770.

Krahé, C., Hahn, U., & Whitney, K. (2013). Is seeing (musical) believing? The eye versus the ear in emotional responses to music. Psychology of Music. DOI: 0305735613498920.

Leman, M. (2008). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746.

Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: from motion to'e'motion. *Social cognitive and affective neuroscience*, 1(3), 235-241.

+REFERENCES (2/2)

Morrison, S. J., Price, H. E., Geiger, C. G., & Cornacchio, R. A. (2009). The effect of conductor expressivity on ensemble performance evaluation. Journal of Research in Music Education, 57, 37–49.

Petrini, K., McAleer, P., & Pollick, F. (2010). Audiovisual integration of emotional signals from music improvisation does not depend on temporal correspondence. Brain Research, 1323, 139–148.

Quinto, L., Thompson, W. F., Russo, F. A., & Trehub, S. E. (2010). A comparison of the McGurk effect for spoken and sung syllables. *Attention, Perception, & Psychophysics*, *72*(6), 1450-1454.

Rosenblum, L. D., & Fowler, C. A. (1991). Audiovisual investigation of the loudness-effort effect for speech and nonspeech events. Journal of Experimental Psychology: Human Perception and Performance, 17(4), 976.

Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. Perception & Psychophysics, 54, 406–416

Shapiro, L. (2007). The embodied cognition research programme. Philosophy Compass, 2(2), 338-346.

Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. Journal of Experimental Psychology: Human Perception and Performance, 35, 1791–1810.

Schutz, M., & Lipscomb, S. D. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. Perception, 36, 888–897 Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences*, 110, 14580–14585. DOI: 10.1073/pnas.1221454110

Thompson, M. R., & Luck, G. (2012). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. Musicae Scientiae, 16(1), 19-40.

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, 156(1), 203-227.

Toiviainen, P., Luck, G. & Thompson, M. (2010). Embodied meter: hierarchical eigenmodes in music-induced movement. Music Perception, 28(1), 59-70.

Verron, C. (2005). Traitement et visualisation de données gesturalles captées par Optotrak [Processing and visualizing gesture data captured by Optotrak]. Unpublished Report. Input Devices and Music Interaction Laboratory, McGill University. Retrieved from http://www.idmil.org/publications

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, *34*(1), 97-113.