

What makes listeners infer teaching intentions?

Atsuko Tominaga<sup>1,\*</sup>, Günther Knoblich<sup>1</sup>, & Natalie Sebanz<sup>1</sup>

<sup>1</sup> Department of Cognitive Science, Central European University, Quellenstraße 51, 1100  
Vienna, Austria

\* Corresponding author: Tominaga\_Atsuko@phd.ceu.edu

## Abstract

Social learning plays an important role in human skill transmission. It has been known that experts tend to change their behaviour particularly for teaching purposes. Our previous research demonstrated that expert pianists systematically modulated their sound so as to teach musical expressive techniques such as articulation and dynamics. For example, pianists played slower and exaggerated each technique when they had an intention to teach. In order to achieve successful skill transmission, it is vital to examine whether such sound modulations are identified and used to infer teaching intentions by novices, who are going to acquire skills. The current experiments investigated which features of piano performance make musicians (in the role of potential learners) infer teaching intentions. Musicians listened to a number of piano recordings where a musical expressive technique of either articulation or dynamics was implemented. Musicians were asked to judge whether each performance was produced in order to teach the designated expressive technique. We quantified recordings with regard to tempo, articulation and dynamics. Overall, slower tempo contributed to musicians' judgments as teaching. Moreover, performances with exaggeration for each technique (e.g., shorter staccato, larger contrast between forte and piano) were more likely to be judged as teaching. These performance features are consistent with expert pianists' sound modulations for teaching purposes. Therefore, musicians (in the role of potential learners) seem to be able to identify didactic sound modulations, only by listening to recorded performances.

*Keywords:* teaching, intention, skill transmission, musical expression

What makes listeners infer teaching intentions?

## Introduction

Learning from others is one of the important elements of skill acquisition. Not only are we able to learn by observing and imitating others, but also we benefit greatly from interacting with others such as teachers and peers (Tomasello, Kruger, & Ratner, 1993). Adults are often being pedagogical to children in order to explain and transmit cultural conventions (Csibra & Gergely, 2009). Active teaching seems to play a crucial role not only to transit skills over generations but also to further develop sophisticated cultures, which cannot be achieved by one single individual or generation (Tennie, Call, & Tomasello, 2009). From a learner's perspective, it is important to identify informative teachers and infer teachers' expectations so that learners can acquire skills through interacting with teachers (Gweon, 2020; Veissière, Constant, Ramstead, Friston, & Kirmayer, 2020).

In pedagogical settings where teachers are supposed to convey useful information to learners, it has been found that teachers often modulate their behaviour for teaching purposes. For example, adults are likely to modulate their speech and action for infants so as to help infants acquire skills (e.g., Brand, Baldwin, & Ashburn, 2002; Kuhl, 2004). Also, some studies have revealed that even towards adult learners, people modulated their speech and action in the similar way as they did for infants (McEllin, Knoblich, & Sebanz, 2017; Uther, Knoll, & Burnham, 2007). These findings suggest that there are general pedagogical behaviours such as slow demonstration and exaggeration of speech and action to send teaching intentions to learners.

Tominaga, Knoblich & Sebanz (*in prep*) extended this line of research to expertise transmission where skills to be acquired are much more complicated than general knowledge, which is mostly acquired throughout development (e.g., Csibra & Gergely, 2009). Expertise usually demands an enormous amount of practice to acquire so that only experts can accomplish a higher level of performance including artistic expression. Based on the

assumption that expert pianists are very good at controlling and modulating their sound to communicate, we investigated how expert pianists could modulate their performance when they were intending to teach musical expressive techniques of articulation (the smoothness of sound) and dynamics (the loudness of sound). The results demonstrated that expert pianists could successfully modulate their performance by playing slower or exaggerating relevant aspects of the performance (e.g., producing shorter staccato or making a larger contrast between forte and piano) to teach musical expressive techniques. Therefore, it seems that even in the domain of expertise, experts exhibit general pedagogical behaviours to communicate with their potential learners.

In order to achieve successful skill transmission, it is necessary to investigate whether novice learners can identify teaching intentions produced by experts. A recent study by McEllin, Sebanz, and Knoblich (2018) demonstrated that people could identify informative intentions such as acting with others or teaching by relying on specific kinematics cues (e.g., velocity profiles of movements). Some research in the field of action expertise suggest that motor simulation (i.e., using one's own motor systems to simulate an action) plays a key role to understand observed actions. For example, expert basketball players were better at predicting the outcome of a free shot compared to expert observers (coaches or sports journalists) and novices (Aglioti, Cesari, Romani, & Urgesi, 2008). Also, it was found that deceptive intentions of fake basketball passes were detected from non-verbal bodily cues only by expert basketball players (Sebanz & Shiffrar, 2009). Taken together, people seem to be able to infer informative intentions by observing actions and motor expertise may enable accurate action prediction.

In the domain of music, musical emotion is often communicated by sound alone and listeners seem to be able to infer performers' intended emotions only by listening to recorded performances (e.g., Akkermans et al., 2019; Gabrielsson & Juslin, 1996). It has been known that musicians are generally more accurate to decode emotions from performances (Battcock

& Schutz, 2021; Lima & Castro, 2011). Repp and Knoblich (2004) also found that expert pianists could distinguish their own performance from others' performances by listening to recordings. These results suggest that musical expertise may be important to accurately identify and infer intentions by listening to recorded performances.

The current experiments investigate which features of piano performance make musicians (in the role of potential learners) infer teaching intentions, only by listening to recorded performances. We collected various piano recordings from our previous experiments (Tominaga et al., *in prep*). The obtained recordings consisted of two pieces (one was used for Experiment 1 and the other was used for Experiment 2 in the current paper). The two pieces were performed by various pianists multiple times, therefore the recordings contained many variations in terms of timing, articulation and dynamics. In the current experiments, we asked musically trained participants to judge which performance was performed for teaching purposes. If teachers' intentions are successfully communicated with learners, pedagogical behavioural features in our previous experiments such as slower demonstration and exaggerated performance should be identified and used to infer teaching intentions. In Experiment 1, we examined which features of piano performance were correlated with participants judgments as "teaching". Experiment 2 was conducted to replicate what we found in Experiment 1 with a more naturalistic piece of music.

## **Experiment 1**

### **Methods**

#### **Participants**

We recruited 21 participants who had at least six years of training in any musical instrument. They were able to read sheet music and knew two musical expressive techniques of articulation and dynamics. One participant was excluded due to an experimental error. Therefore, 20 participants (Female: 13) were included for data analysis and had 11.85 years

of musical training on average ( $SD = 5.62$ ). They were all right-handed with a mean age of 28.80 ( $SD = 9.09$ ). All participants were recruited through an online participant platform (SONA system, <https://www.sona-systems.com>). The study (No. 2020/02) was approved by the Psychological Research Ethics Board (PREBO) CEU PU in Austria.

## Apparatus

The experiment was programmed in Python 3.8.2 using the PsychoPy Python library (2020.2.4; <https://www.psychopy.org/>) on a Mac Book Pro with Mac OS X Catalina 10.15.6. Stimuli were played using the Mido Python library (1.2.9; <https://mido.readthedocs.io/en/latest/>) on a Max/MSP patcher (8.1.7; <https://cycling74.com/products/max>). During the experiment, participants listened to the stimuli via headphones (Audio-Technica ATH-M50X).

## Stimuli

We selected stimuli from our previous experiments (Tominaga et al., *submitted*). Stimuli were produced by actual pianists on a weighed Yamaha MIDI (Musical Instrument Digital Interface) digital piano and recorded as MIDI files. Multiple pianists played one piece of music with a musical expressive techniques of either articulation (*Figure 1, A*) or dynamics (*Figure 1, B*). Articulation refers to the smoothness of sound, which is comprised of legato and staccato. Legato indicates smooth and connected sound whereas staccato indicates sharp and separate sound. Dynamics refers to the loudness of sound, which is comprised of forte and piano. Forte indicates loud sound while piano indicates soft sound. The piece was taken from “A Dozen a Day - Play with Ease in Many Keys” by Edna-Mae Burnam and modified for the experiment. The stimuli were performed around 80 quarter-beats per minute.

In Tominaga et al., (*submitted*), participants were asked to perform the piece with either articulation or dynamics in two different conditions. In the teaching condition, participants were instructed to perform the piece with the designated expressive technique as

if they were teaching it to students (e.g., in a lesson). In the performing condition, participants were instructed to perform the piece with the designated expressive technique as if they were performing it to an audience (e.g., in a concert). In total, we obtained 453 valid performances (i.e., performances without any pitch error) from the teaching condition and 436 valid performances from the performing condition.

For the current experiment, 96 performances were chosen from the valid performances in Tominaga et al. (*submitted*; Experiment 1). We randomly sampled 24 articulation performances and 24 dynamics performances from the teaching condition as well as 24 articulation performances and 24 dynamics performances from the performing condition. It is important to note that the performances from the teaching condition did not necessarily exhibit specific features of teaching that we found in our previous experiments (e.g., exaggeration) since we randomly sampled the performances.

## Procedure

Upon arrival, participants read information sheet about the experiment and gave informed consent prior to participation. In the experiment, all instructions were displayed on a computer screen in front of the participants and an experimenter also explained the procedure. Participants were instructed that they were going to listen to piano recordings with musical expression (either articulation or dynamics), which were either produced as if a pianist were teaching the expression to students (e.g., in a lesson; teaching) or as if a pianist were performing it to an audience (e.g., in a concert; performing). In each trial, participants listened to one recording and were asked whether the recording was produced for teaching purposes or not. Participants responded by pressing either a yes (for teaching; left arrow key) or no (for performing; right arrow key) button. While listening to each recording, sheet music, which corresponded to the recording, was shown on the screen in front of the participants (*Figure 2*).

There were two blocks and each block only included recordings with one musical expressive technique of either articulation or dynamics. Each block consisted of 4 practice trials and 48 experimental trials. Each recording was evaluated only once in the experiment. The order of the blocks was counterbalanced across the participants. The order of the recordings was randomised within each block.

At the end of the experiment, participants filled in a questionnaire about their demographic information and experience in musical instruments.

### Data analysis

Data processing and statistical analysis were performed in R version 4.0.5. Correlation analysis was performed with the standard *cor* function and regression models for multiple regression were fit with the standard *lm* function from the *stats* R package. Stimuli (MIDI files) were converted to numerical data in terms of time, pitch and velocity for the onset and offset of each note using the *tuneR* R package (<https://cran.r-project.org/web/packages/tuneR/tuneR.pdf>).

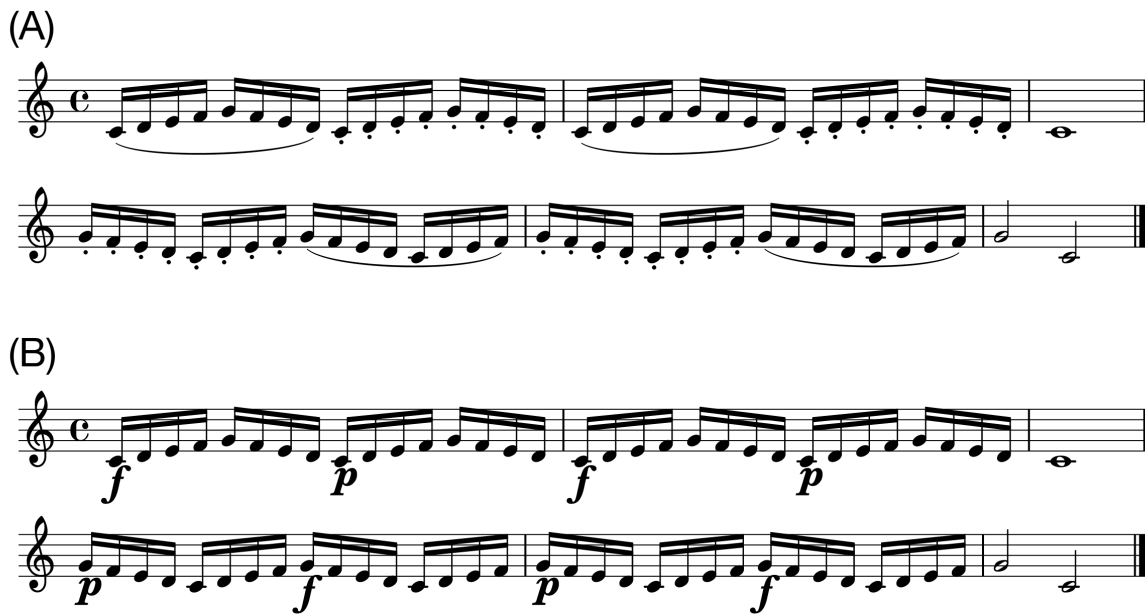
Stimuli were quantified with regard to tempo (interonset intervals; IOIs), articulation (key-overlap time; KOT), dynamics (key velocity; KV) and dynamics contrast (key velocity difference; KV-Diff) only for 16th notes. Interonset intervals are time intervals between onsets of adjacent notes. Larger IOIs indicate slower tempo while smaller IOIs indicate faster tempo. Key-overlap time is the time overlap between two adjacent notes, namely the difference between the offset time of the current note and the onset time of the ensuing note (e.g., Bresin & Battel, 2000). Positive KOT values indicate legato styles whereas negative KOT values indicate staccato styles. Key velocity is obtained from MIDI data to describe how fast a performer hit the key. Larger KV values indicate forte styles while smaller KV values indicate piano styles. Additionally, we also measured dynamics contrast where one subcomponent of the technique moves to the other (e.g., from forte to piano) to illustrate



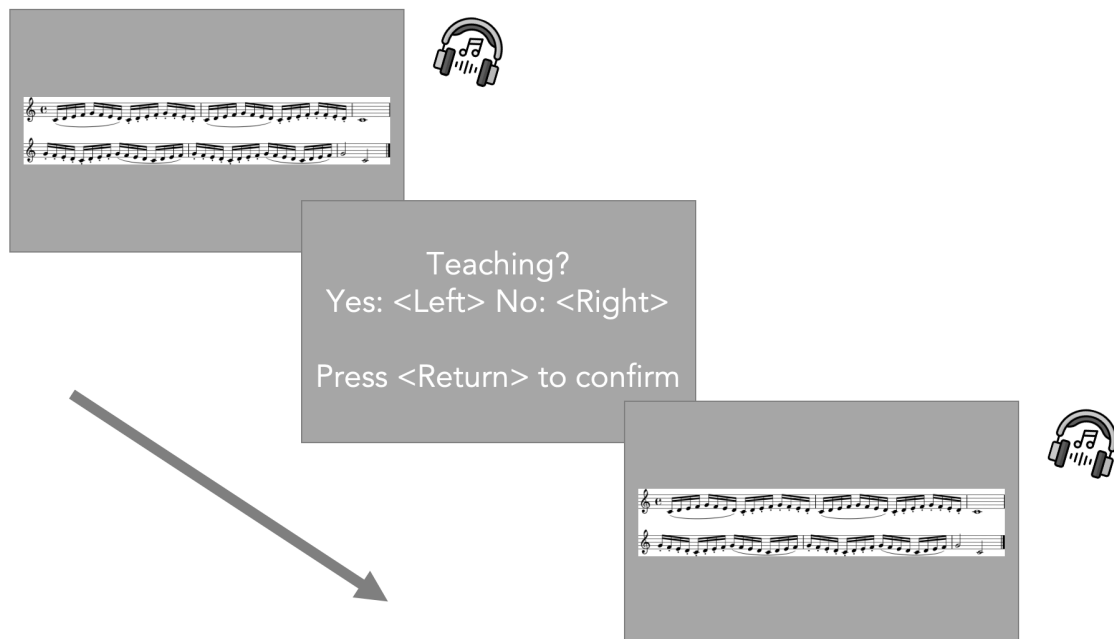
how much dynamics contrast a performer made at transition points.

First, we investigated data by looking at correlations between performance features (i.e., IOIs, KOT, KV, KV-Diff) and participants' judgments as "teaching" (i.e., what percentage of participants responded as "yes") for each stimulus. Second, we run multiple regression and entered all four performance features to examine which predictor significantly contributed to participants' judgments as teaching. Since articulation and dynamics were comprised of two opposite directional values (i.e., legato vs. staccato, forte vs. piano), we entered each subcomponent (i.e., legato, staccato, forte, piano) separately to represent either an articulation or dynamics feature of the performance for a regression model.

For articulation recordings, there were two models. The first model considered only legato parts of the stimuli for KOT and KV, and transition points from legato to staccato for KV-Diff. The second model considered only staccato parts of the stimuli for KOT and KV, and transition points from staccato to legato for KV-Diff. For dynamics recordings, there were two models. The first model considered only forte parts of the stimuli for KV and KOT, and transition points from forte to piano for KV-Diff. The second model considered only piano parts of the stimuli for KV and KOT, and transition points from piano to forte for KV-Diff. With regard to tempo (IOIs), there was only one value for each technique because tempo was consistent across the performance and was not dependent on subcomponents. Therefore, we entered the same tempo value for legato and staccato or forte and piano parts.



*Figure 1.* Stimuli. (A) Articulation. The curved line (slur) indicates legato and the dots indicate staccato. (B) Dynamics. The symbol ‘f’ denotes forte and the symbol ‘p’ denotes piano. Only the 16th notes were used for data analysis.



*Figure 2.* Procedure. Participants listened to a recording via headphones while corresponding sheet music was displayed on a monitor. They were required to respond by pressing a left-arrow (yes) or right-arrow (no) key for each judgment.

## Results

All results were reported as significant at  $p < 0.05$ .

### Correlation

A Pearson product-moment correlation coefficient was computed by default for correlation analysis and a Spearman's rank correlation coefficient was additionally computed if normality assumption was violated based on the Shapiro-Wilk normality test.

**Tempo (IOIs).** Performance tempi (IOIs) were significantly correlated with participants' judgments as teaching for both techniques (Articulation;  $r = .77$ ,  $p < 0.001$ , Dynamics;  $r = .42$ ,  $p = 0.003$ , *Figure 3*). Participants tended to identify slower performances as teaching.

However, the Shapiro-Wilk normality test revealed that the distribution of IOIs for dynamics recordings was not normally distributed ( $p = 0.037$ ). Therefore, a Spearman's rank correlation coefficient was additionally used to assess the relationship between performance tempi of dynamics recordings and participants' judgments. The result failed to reach significance ( $r = .27$ ,  $p = 0.066$ ), suggesting that we need to be careful about the interpretation of the relationship between performance tempi of dynamics recordings and participants' judgments as teaching.

**Articulation (KOT).** For articulation recordings, there was a significant relationship between KOT values and participants' judgments as teaching (*Figure 4*, left). Specifically, performances with shorter staccato ( $r = -.73$ ,  $p < 0.001$ ) and longer legato ( $r = .40$ ,  $p = 0.005$ ) were more likely to be judged as teaching.

For dynamics recordings, there was no significant relationship between KOT values and participants' judgments as teaching (Forte;  $r = -.04$ ,  $p = 0.81$ , Piano;  $r = -.19$ ,  $p = 0.19$ , *Figure 4*, right).

The Shapiro-Wilk normality test revealed that the distribution of KOT for dynamics

recordings was not normally distributed (Forte;  $p < 0.001$ , Piano;  $p < 0.001$ ). Spearman's rank correlation coefficients also showed that there was no significant relationship between KOT values and participants' judgments as teaching (Forte;  $r = .15$ ,  $p = 0.31$ , Piano;  $r = -.12$ ,  $p = 0.43$ ).

**Dynamics (KV).** For dynamics recordings, there was a significant relationship between KV values and participants' judgments as teaching (*Figure 5*, right). Specifically, performances with louder forte were more likely to be judged as teaching ( $r = .45$ ,  $p = 0.001$ ). However, there was no significant relationship between KV values for piano and participants' judgments as teaching ( $r = -.22$ ,  $p = 0.13$ ).

For articulation recordings, there was no significant relationship between KV values and participants' judgments as teaching (Legato;  $r = .10$ ,  $p = 0.52$ , Staccato;  $r = -.02$ ,  $p = 0.87$ , *Figure 5*, left).

The Shapiro-Wilk normality test revealed that the distribution of KOT for articulation recordings for legato was not normally distributed ( $p = 0.002$ ). A Spearman's rank correlation coefficients also showed that there was no significant relationship between KOT values for legato and participants' judgments as teaching ( $r = .03$ ,  $p = 0.84$ ).

**Dynamics contrast (KV-Diff).** For dynamics recordings, there was a significant relationship between KV difference between forte and piano and participants' judgments as teaching (*Figure 6*, right). Specifically, performances with larger contrasts between forte and piano were more likely to be judged as teaching (From Forte to Piano;  $r = -.50$ ,  $p < 0.001$ , From Piano to Forte;  $r = .62$ ,  $p < 0.001$ ).

For articulation recordings, there was no significant relationship between KV difference between legato and staccato and participants' judgments as teaching (From Legato to Staccato;  $r = -.26$ ,  $p = 0.071$ , From Staccato to Legato;  $r = -.19$ ,  $p = 0.21$ , *Figure 6*, left).

The Shapiro-Wilk normality test revealed that the distribution of KV difference for

articulation recordings for transition points from legato to staccato was not normally distributed ( $p = 0.007$ ). A Spearman's rank correlation coefficients also showed that there was no significant relationship between KOT values for legato and participants' judgments as teaching ( $r = -.23$ ,  $p = 0.12$ ).

## Multiple regression

In order to further explore which feature of performance contributed the most to participants' judgments as teaching, multiple regression analyses were conducted. Statistical model assumptions were tested using the *performance* R package (Lüdtke, Ben-Shachar, Patil, Waggoner, & Makowski, 2021) and all assumptions were met. Since articulation and dynamics consisted of two opposite subcomponents (i.e., legato vs. staccato, forte vs. piano) and therefore cannot be summed up to represent each technique as one value, we reported four separate regression models for each subcomponent (see details in *Data analysis*).

**Legato.** A multiple regression analysis was conducted to predict participants' judgments as teaching based on performance features of tempo (IOIs), articulation (KOT for legato parts), dynamics (KV for legato parts) and dynamics contrast (KV-Diff from legato to staccato). The result of the regression indicated that the model explained 64.63 % of the variance ( $F(4, 43) = 22.50$ ,  $p < 0.001$ ). It was found that tempo (IOIs;  $\beta = 0.78$ ,  $p < 0.001$ ) and articulation for the legato parts (KOT;  $\beta = 0.26$ ,  $p = 0.004$ ) were significant predictors of participants' judgments as teaching.

**Staccato.** A multiple regression analysis was conducted to predict participants' judgments as teaching based on performance features of tempo (IOIs), articulation (KOT for staccato parts), dynamics (KV for staccato parts) and dynamics contrast (KV-Diff from staccato to legato). The result of the regression indicated that the model explained 63.97 % of the variance ( $F(4, 43) = 21.90$ ,  $p < 0.001$ ). It was found that tempo (IOIs;  $\beta = 0.52$ ,  $p = 0.002$ ) and articulation for the staccato parts (KOT;  $\beta = -0.28$ ,  $p = 0.020$ ) were significant predictors of participants' judgments as teaching.

**Forte.** A multiple regression analysis was conducted to predict participants' judgments as teaching based on performance features of tempo (IOIs), articulation (KOT for forte parts), dynamics (KV for forte parts) and dynamics contrast (KV-Diff from forte to piano). The result of the regression indicated that the model explained 35.90 % of the variance ( $F(4, 43) = 7.60, p < 0.001$ ). It was found that tempo (IOIs;  $\beta = 0.35, p = 0.007$ ) and dynamics for the forte parts (KV;  $\beta = 0.73, p = 0.048$ ) were significant predictors of participants' judgments as teaching.

**Piano.** A multiple regression analysis was conducted to predict participants' judgments as teaching based on performance features of tempo (IOIs), articulation (KOT for piano parts), dynamics (KV for piano parts) and dynamics contrast (KV-Diff from piano to forte). The result of the regression indicated that the model explained 51.21 % of the variance ( $F(4, 43) = 13.30, p < 0.001$ ). It was found that tempo (IOIs;  $\beta = 0.38, p < 0.001$ ) and dynamics contrast from piano to forte (KV-Diff;  $\beta = 0.99, p < 0.001$ ) were significant predictors of participants' judgments as teaching.

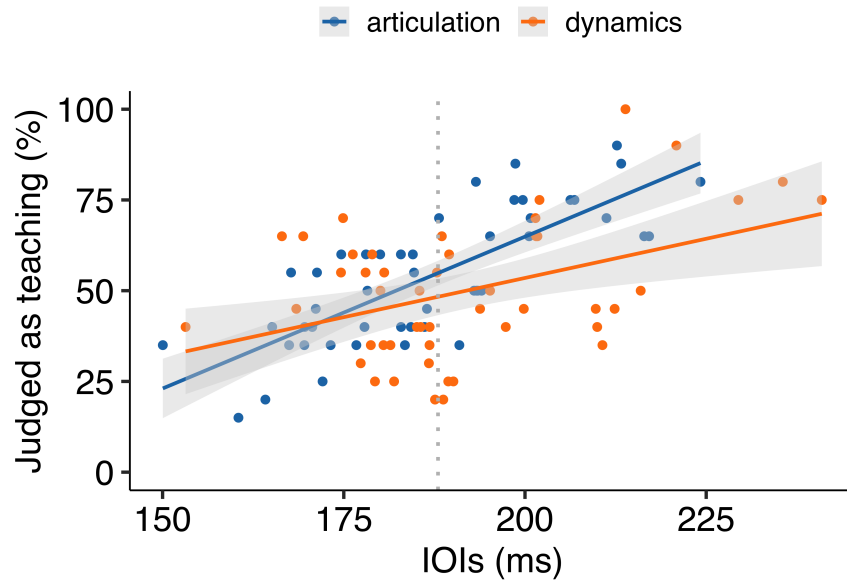


Figure 3. Experiment 1: Scatter plot showing the correlation between tempo feature (IOIs) and average participants' judgments as teaching for each recording. Therefore, each dot represents each stimulus.

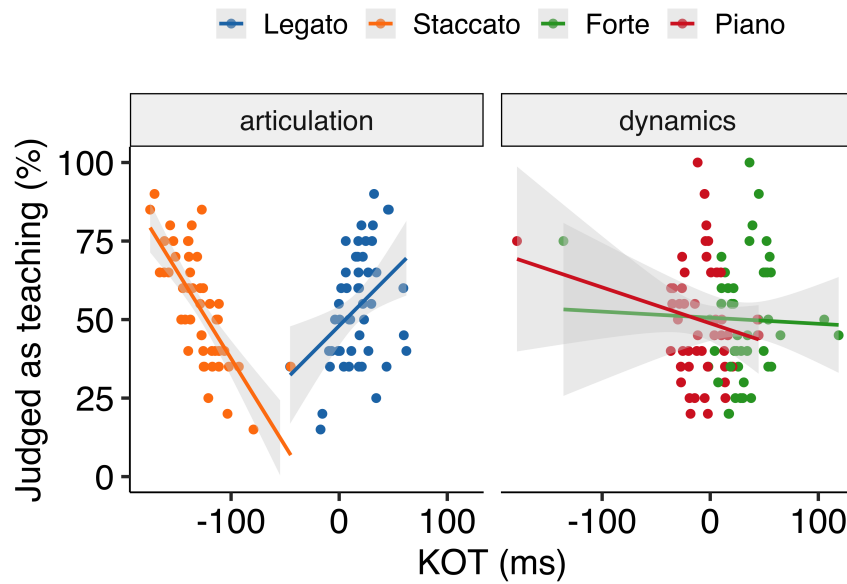


Figure 4. Experiment 1: Scatter plot showing the correlation between articulation feature (KOT) and average participants' judgments as teaching for each recording. Therefore, each dot represents each stimulus.



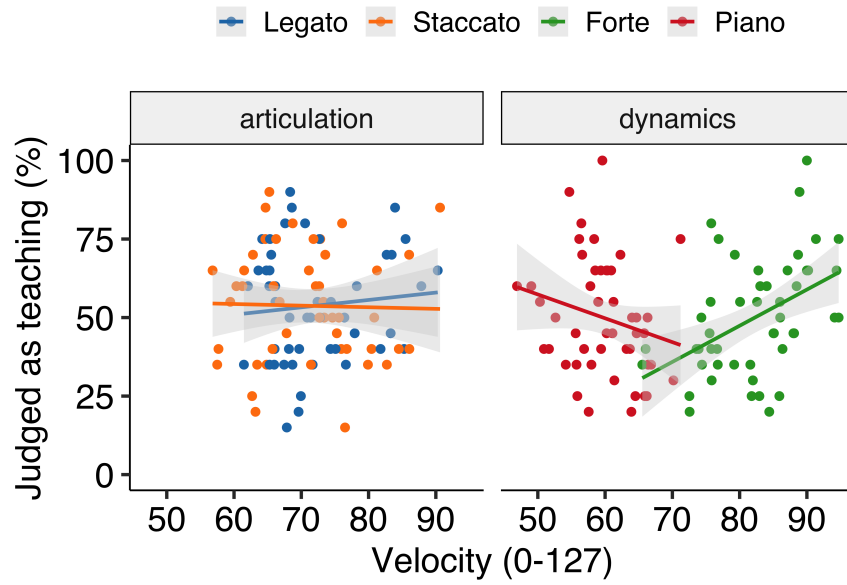


Figure 5. Experiment 1: Scatter plot showing the correlation between dynamics feature (KV) and average participants' judgments as teaching for each recording. Therefore, each dot represents each stimulus.

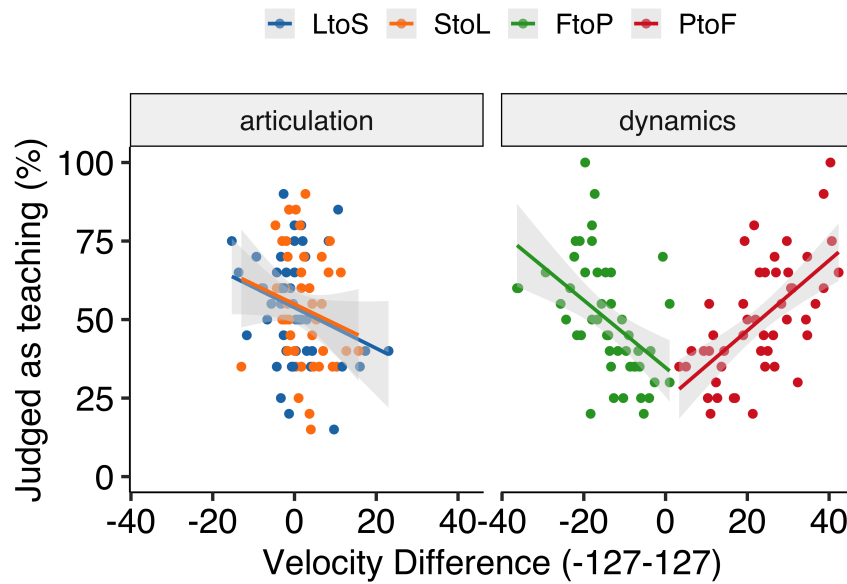


Figure 6. Experiment 1: Scatter plot showing the correlation between dynamics contrast feature (KV-Diff) and average participants' judgments as teaching for each recording. Therefore, each dot represents each stimulus.

## Discussion

Experiment 1 investigated which features of piano performance made musicians (in the role of potential learners) infer teaching intentions. The results demonstrated that performances with slower tempo were more likely to be judged as teaching by musicians regardless which expressive technique was implemented in the piece. For articulation recordings, performances with longer legato and shorter staccato were tended to be judged as teaching. For dynamics recordings, performances with louder forte were more likely to be judged as teaching whereas there was no relationship between softer sound (i.e., piano) and participants' judgments as teaching. Importantly, performances with larger contrasts between forte and piano for both directions (i.e., from forte to piano, from piano to forte) were more likely to be judged as teaching. This result may suggest that dynamics contrast might be reliably used to infer teaching intentions, rather than absolute dynamics values themselves. Moreover, multiple regression analyses implied that tempo feature is the strongest predictor of participants' judgments as teaching in general whereas there were specific predictors depending on which expressive technique was implemented in the piece. These performance features were overall consistent with what expert pianists did in our previous experiments for teaching purposes. Therefore, our findings suggest that musicians may rely on generic pedagogical behaviours (e.g., slower demonstration, exaggeration) to infer teaching intentions of expert pianists, only by listening to recorded performances.

## Experiment 2

The aim of Experiment 2 was to replicate our previous findings in Experiment 1 with a more naturalistic piece of music. Given the findings on Experiment 1, we predicted that slower performance would be likely to be judged as teaching. Also performances with exaggerated articulation and dynamics (in particular, longer legato and shorter staccato, larger contrast between forte and piano) would be likely to be judged as teaching.

## Methods

### Participants

We recruited 21 participants who had at least six years of training in any musical instrument or singing. They were able to read sheet music and know two musical expressive techniques of articulation and dynamics. One participant was excluded because s/he did not understand the instructions. Therefore, 20 participants (Female: 10) were included for data analysis and had 12.65 years of training on average in any musical instrument or singing ( $SD = 5.40$ ). Most people were right-handed (Left: 1) with a mean age of 33.55 ( $SD = 12.80$ ). As Experiment 1, all participants were recruited through the SONA system and the study (No. 2020/02) was approved by the PREBO CEU PU in Austria.

### Apparatus and procedure

The apparatus and procedure were identical to Experiment 1 except that each block consisted of 4 practice trials and 36 experimental trials. The number of trials was reduced due to the time constraint of the experiment.

### Stimuli

As Experiment 1, we selected stimuli from our previous experiments (Tominaga et al., *submitted*). The excerpt was taken from “Sonatina Op.36 (No.3) in C major” by Muzio Clementi and modified for the experiment. The excerpt was performed with either

articulation (*Figure 7, A*) or dynamics (*Figure 7, B*). The stimuli were performed around 100 - 120 quarter-beats per minute (i.e., around 250 - 300 ms for 8th-note IOIs).

For the current experiment, 96 performances were chosen from the valid performances in Tominaga et al. (*submitted*; Experiment 2). There were 248 valid performances in the teaching condition and 256 valid performances in the performing condition. We randomly sampled 18 articulation performances and 18 dynamics performances from the teaching condition as well as 18 articulation performances and 18 dynamics performances from the performing condition. In total, 72 recordings were selected as stimuli for the current experiment. Again, it is important to note that each performance from the teaching condition did not necessarily exhibit specific features of teaching that we found in the previous experiments (e.g., exaggeration) since we randomly sampled the performances.

## Data analysis

The data analysis was identical to Experiment 1. Only the 8th notes with expressive notations were included for data analysis. As a result, only one 8th note in the 4th measure without any expression was not included.

## Results

All results were reported as significant at  $p < 0.05$ .

## Correlation

As Experiment 1, a Pearson product-moment correlation coefficient was computed by default for correlation analysis and a Spearman's rank correlation coefficient was additionally computed if normality assumption was violated based on the Shapiro-Wilk normality test.

**Tempo (IOIs).** Unlike Experiment 1, there was no significant relationship between performance tempi (IOIs) and participants' judgments as teaching for both techniques (Articulation;  $r = .25$ ,  $p = 0.147$ , Dynamics;  $r = .39$ ,  $p = 0.019$ , *Figure 8*).



Figure 7. Stimuli. (A) Articulation. The curved line (slur) indicates legato and the dots indicate staccato. (B) Dynamics. The symbol ‘f’ denotes forte and the symbol ‘p’ denotes piano. Only the 8th notes with expressive notations were used for data analysis.

The Shapiro-Wilk normality test revealed that the distribution of IOIs for dynamics recordings was not normally distributed ( $p = 0.080$ ). A Spearman's rank correlation coefficient also showed that there was no significant relationship between tempi for dynamics performances and participants' judgments as teaching ( $r = .29$ ,  $p = 0.091$ ).

**Articulation (KOT).** Unlike Experiment 1, for articulation recordings, there was no significant relationship between KOT values and participants' judgments as teaching (Legato;  $r = -.15$ ,  $p = 0.390$ , Staccato;  $r = -.03$ ,  $p = 0.880$ , *Figure 9*, left).

For dynamics recordings, there was no significant relationship between KOT values for forte and participants' judgments as teaching ( $r = -.11$ ,  $p = 0.52$ ). However, there was a significant relationship between KOT values for piano and participants' judgments as teaching ( $r = -.35$ ,  $p = 0.03$ ), suggesting that performances with louder staccato were more likely to be considered as teaching performance (*Figure 9*, right).

**Dynamics (KV).** As Experiment 1, for dynamics recordings, there was a significant relationship between KV values and participants' judgments as teaching (*Figure 10*, right). Specifically, performances with louder forte ( $r = .45$ ,  $p = 0.007$ ) and softer piano ( $r = -.45$ ,  $p = 0.01$ ) were more likely to be judged as teaching.

The Shapiro-Wilk normality test revealed that the distribution of KV values for dynamics recordings was not normally distributed (Forte;  $p = 0.048$ , Piano;  $p = 0.004$ ). Spearman's rank correlation coefficient also confirmed that performances with louder forte ( $r = .43$ ,  $p = 0.010$ ) and softer piano ( $r = -.45$ ,  $p = 0.01$ ) were more likely to be judged as teaching.

For articulation recordings, there was no significant relationship between KV values and participants' judgments as teaching (Legato;  $r = .08$ ,  $p = 0.63$ , Staccato;  $r = .21$ ,  $p = 0.22$ , *Figure 10*, left).

**Dynamics contrast (KV-Diff).** As Experiment 1, for dynamics recordings, there was a significant relationship between KV difference between forte and piano and

participants' judgments as teaching (*Figure 11*, right). Specifically, performances with larger contrasts between forte and piano were more likely to be judged as teaching (From Forte to Piano;  $r = -.75$ ,  $p < 0.001$ , From Piano to Forte;  $r = .59$ ,  $p < 0.001$ ).

The Shapiro-Wilk normality test revealed that the distribution of KV-Diff values for dynamics recordings was not normally distributed (Forte to Piano;  $p = 0.850$ , Piano to Forte;  $p = 0.008$ ). Spearman's rank correlation coefficient also confirmed that performances with larger contrasts between forte and piano were more likely to be judged as teaching (Forte to Piano;  $r = -.73$ ,  $p < 0.001$ , Piano to Forte;  $r = .59$ ,  $p < 0.001$ ).

### Multiple regression

As Experiment 1, we performed multiple regression analyses to further explore which feature of performance contributed the most to participants' judgments as teaching. Again, since articulation and dynamics consisted of two opposite subcomponents (i.e., legato vs. staccato, forte vs. piano) and therefore cannot be summed up to represent each technique as one value, we reported four separate regression models for each subcomponent (see details in *Data analysis* in Experiment 1).

**Legato.** A multiple regression analysis was conducted to predict participants' judgments as teaching based on performance features of tempo (IOIs), articulation (KOT for legato parts), dynamics (KV for legato parts) and dynamics contrast (KV-Diff from legato to staccato). The result of the regression indicated that the model explained 19.77 % of the variance ( $F(4, 31) = 3.20$ ,  $p = 0.03$ ). It was found that tempo (IOIs;  $\beta = 0.58$ ,  $p = 0.011$ ), dynamics (KV) and dynamics contrast (KV-Diff) for the legato parts were significant predictors of participants' judgments as teaching. However, articulation (KOT;  $\beta = -0.08$ ,  $p = 0.466$ ) for the legato parts was not a significant predictor as opposed to Experiment 1.

**Staccato.** A multiple regression analysis was conducted to predict participants' judgments as teaching based on performance features of tempo (IOIs), articulation (KOT for

staccato parts), dynamics (KV for staccato parts) and dynamics contrast (KV-Diff from staccato to legato). The overall was not statistically significant ( $R^2 = 12.28$ ,  $F(4, 31) = 2.20$ ,  $p = 0.09$ ).

**Forte.** A multiple regression analysis was conducted to predict participants' judgments as teaching based on performance features of tempo (IOIs), articulation (KOT for forte parts), dynamics (KV for forte parts) and dynamics contrast (KV-Diff from forte to piano). The result of the regression indicated that the model explained 60.69 % of the variance ( $F(4, 31) = 14.50$ ,  $p < 0.001$ ). It was found that dynamics contrast from forte to piano (KV-Diff;  $\beta = -1.64$ ,  $p < 0.001$ ) were significant predictors of participants' judgments as teaching.

**Piano.** A multiple regression analysis was conducted to predict participants' judgments as teaching based on performance features of tempo (IOIs), articulation (KOT for piano parts), dynamics (KV for piano parts) and dynamics contrast (KV-Diff from piano to forte). The result of the regression indicated that the model explained 49.49 % of the variance ( $F(4, 31) = 9.60$ ,  $p < 0.001$ ). It was found that tempo (IOIs;  $\beta = 0.55$ ,  $p = 0.022$ ) and dynamics contrast from piano to forte (KV-Diff;  $\beta = 1.09$ ,  $p < 0.001$ ) were significant predictors of participants' judgments as teaching.

## Discussion

Experiment 2 aimed to replicate our previous findings in Experiment 1 with a more naturalistic piece of music. We successfully replicated the findings from Experiment 1 in dynamics features. Performances with louder forte and softer piano were likely to be judged as teaching for dynamics recordings. Moreover, there seems to be a strong relationship between dynamics contrast between forte and piano and participants' judgments as teaching.

However, we failed to replicate the results in terms of tempo (IOIs) and articulation (KOT). Compared with Experiment 1, the piece used in Experiment 2 was musically much complex. Therefore, articulation features may not be that reliable in order to infer teaching



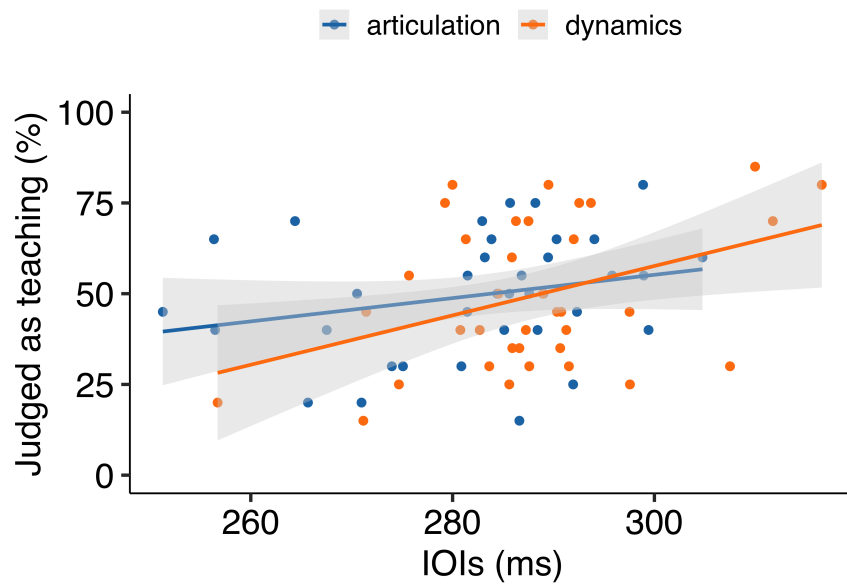


Figure 8. Experiment 2: Scatter plot showing the correlation between tempo features (IOIs) and average participants' judgments as teaching for each recording. Therefore, each dot represents each stimulus.

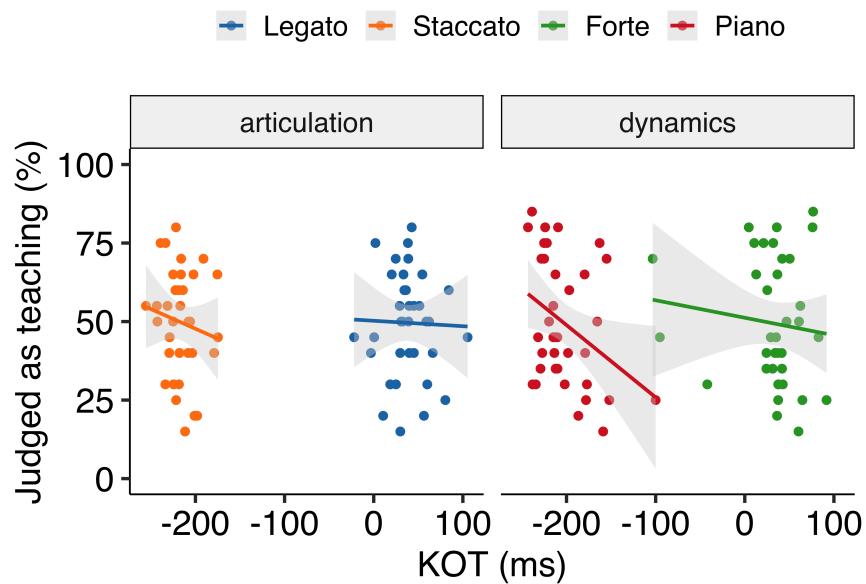


Figure 9. Experiment 2: Scatter plot showing the correlation between articulation features (KOT) and average participants' judgments as teaching for each recording. Therefore, each dot represents each stimulus.

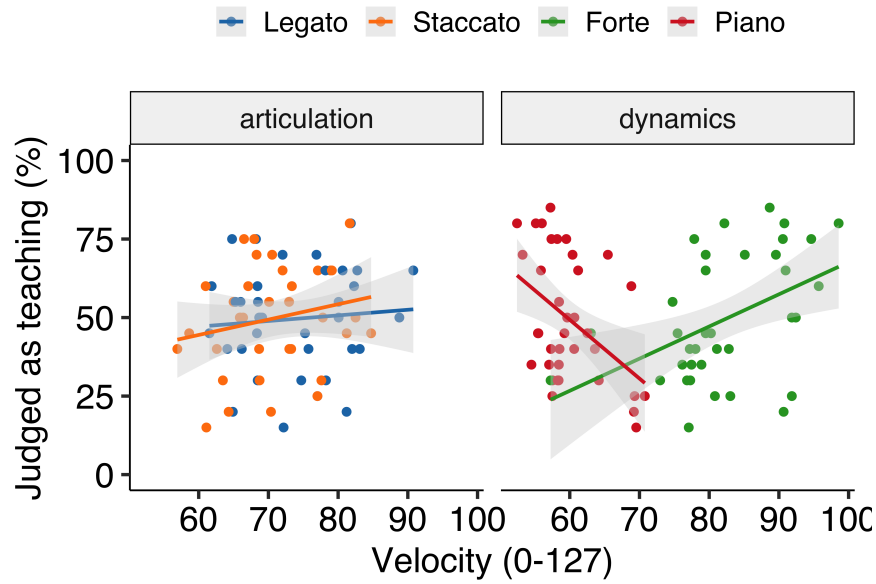


Figure 10. Experiment 2: Scatter plot showing the correlation between dynamics features (KV) and average participants' judgments as teaching for each recording. Therefore, each dot represents each stimulus.

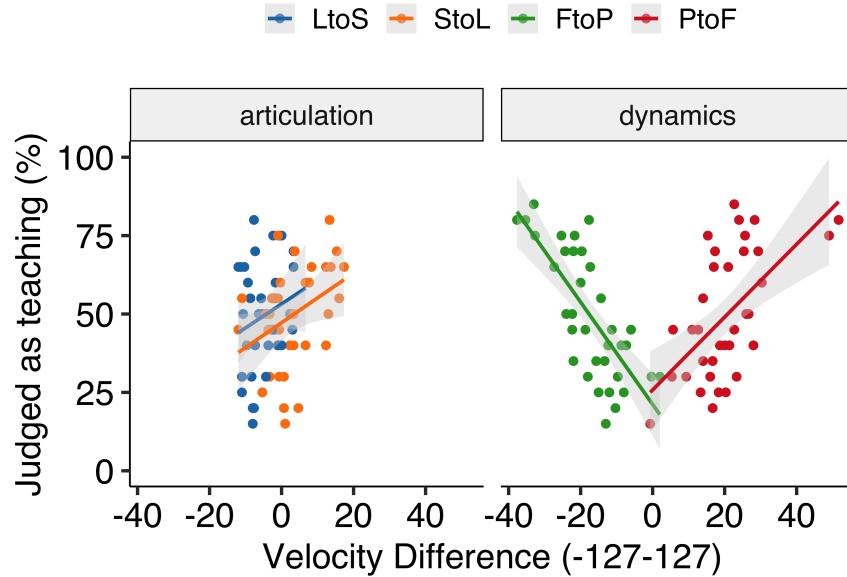


Figure 11. Experiment 2: Scatter plot showing the correlation between dynamics contrast features (KV-Diff) and average participants' judgment as teaching for each stimulus.

intentions. Unlike tempo and loudness, which are relatively easier to understand, articulation is very particular to music. The reason we could not see a relationship between articulation values and participants' judgments as teaching might be coming from conceptual and perceptual difficulties of articulation, even for musicians.

## References

- Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*(9), 1109–1116.  
<https://doi.org/10.1038/nn.2182>
- Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., ... Frieler, K. (2019). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion*, *33*(6), 1099–1118.  
<https://doi.org/10.1080/02699931.2018.1541312>
- Battcock, A., & Schutz, M. (2021). Emotion and expertise: How listeners with formal music training use cues to perceive emotion. *Psychological Research*.  
<https://doi.org/10.1007/s00426-020-01467-1>
- Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for “motionese”: Modifications in mothers’ infant-directed action. *Developmental Science*, *5*(1), 72–83.  
<https://doi.org/10.1111/1467-7687.00211>
- Bresin, R., & Battel, G. U. (2000). Articulation Strategies in Expressive Piano Performance Analysis of Legato, Staccato, and Repeated Notes in Performances of the Andante Movement of Mozart’s Sonata in G Major (K 545). *Journal of New Music Research*, *29*(3), 211–224. <https://doi.org/10.1076/jnmr.29.3.211.3092>
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. <https://doi.org/10.1016/j.tics.2009.01.005>
- Gabrielsson, A., & Juslin, P. N. (1996). Emotional Expression in Music Performance: Between the Performer’s Intention and the Listener’s Experience. *Psychology of Music*, *24*(1), 68–91. <https://doi.org/10.1177/0305735696241007>
- Gweon, H. (2020). The role of communication in acquisition, curation, and transmission of culture. *Behavioral and Brain Sciences*, *43*. <https://doi.org/10.1017/S0140525X19002863>
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. <https://doi.org/10.1038/nrn1533>

- Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion, 11*(5), 1021–1031.  
<https://doi.org/10.1037/a0024521>
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software, 6*(60), 3139.  
<https://doi.org/10.21105/joss.03139>
- McEllin, L., Knoblich, G., & Sebanz, N. (2017). Distinct kinematic markers of demonstration and joint action coordination? Evidence from virtual xylophone playing. *Journal of Experimental Psychology: Human Perception and Performance, 44*(6), 885.  
<https://doi.org/10.1037/xhp0000505>
- McEllin, L., Sebanz, N., & Knoblich, G. (2018). Identifying others' informative intentions from movement kinematics. *Cognition, 180*, 246–258.  
<https://doi.org/10.1016/j.cognition.2018.08.001>
- Repp, B. H., & Knoblich, G. (2004). Perceiving Action Identity: How Pianists Recognize Their Own Performances. *Psychol Sci, 15*(9), 604–609.  
<https://doi.org/10.1111/j.0956-7976.2004.00727.x>
- Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic Bulletin and Review, 16*(1), 170–175.  
<https://doi.org/10.3758/PBR.16.1.170>
- Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1528), 2405–2415. <https://doi.org/10.1098/rstb.2009.0052>
- Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences, 16*(3), 495–511. <https://doi.org/10.1017/S0140525X0003123X>
- Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication, 49*(1), 2–7.

<https://doi.org/10.1016/j.specom.2006.10.003>

Veissière, S. P. L., Constant, A., Ramstead, M. J. D., Friston, K. J., & Kirmayer, L. J.

(2020). Thinking through other minds: A variational approach to cognition and culture.

*Behavioral and Brain Sciences*, 43. <https://doi.org/10.1017/S0140525X19001213>