

Relations between affective music and speech: evidence from dynamics of affective piano performance and speech production

Xiaoluan Liu* and Yi Xu

Department of Speech, Hearing and Phonetic Sciences, University College London, London, UK

OPEN ACCESS

Edited by:

Petri Laukka,
Stockholm University, Sweden

Reviewed by:

Bruno Gingras,
University of Vienna, Austria
Steven Robert Livingstone,
Ryerson University, Canada

*Correspondence:

Xiaoluan Liu,
Department of Speech, Hearing and
Phonetic Sciences, University College
London, Chandler House, 2 Wakefield
Street, London WC1N 1PF, UK
lx0803@outlook.com

Specialty section:

This article was submitted to
Emotion Science,
a section of the journal
Frontiers in Psychology

Received: 21 January 2015

Accepted: 15 June 2015

Published: 08 July 2015

Citation:

Liu X and Xu Y (2015) Relations
between affective music and speech:
evidence from dynamics of affective
piano performance and speech
production. *Front. Psychol.* 6:886.
doi: 10.3389/fpsyg.2015.00886

This study compares affective piano performance with speech production from the perspective of dynamics: unlike previous research, this study uses finger force and articulatory effort as indexes reflecting the dynamics of affective piano performance and speech production respectively. Moreover, for the first time physical constraints such as piano fingerings and speech articulatory constraints are included due to their potential contribution to different patterns of dynamics. A piano performance experiment and speech production experiment were conducted in four emotions: anger, fear, happiness and sadness. The results show that in both piano performance and speech production, anger and happiness generally have high dynamics while sadness has the lowest dynamics. Fingerings interact with fear in the piano experiment and articulatory constraints interact with anger in the speech experiment, i.e., large physical constraints produce significantly higher dynamics than small physical constraints in piano performance under the condition of fear and in speech production under the condition of anger. Using production experiments, this study firstly supports previous perception studies on relations between affective music and speech. Moreover, this is the first study to show quantitative evidence for the importance of considering motor aspects such as dynamics in comparing music performance and speech production in which motor mechanisms play a crucial role.

Keywords: dynamics, emotion, piano performance, speech production, fingerings, articulatory constraints

Introduction

Background

Music and speech reflect fundamental aspects of human capacities (Juslin and Laukka, 2003; Patel, 2008). The parallels between music and speech have been attracting scholarly interest for a long period (Fonagy and Magdics, 1963; Sundberg, 1982; Scherer, 1995), with attempts to compare the two from a wide range of perspectives: prosody (Scherer, 1995), semantics (Seifert et al., 2013), syntax (Lerdahl, 2013), evolution (Cross et al., 2013), neurocognitive mechanisms (Steinbeis and Koelsch, 2008), and facial expressions (Carlo and Guaitella, 2004; Livingstone et al., in press). Particularly, an increasing amount of attention has been given to using perceptual tests for acoustic comparisons between affective music and speech, as they are two important means of emotion communication (Buck, 1984; Wilson, 1994; Juslin and Laukka, 2003) which is crucial for

maintaining social bonds in human society (Ekman, 1992). The majority of comparative studies show that perceptually, acoustic cues (pitch, intensity, and duration) of affective music and speech are similar (Juslin and Laukka, 2003; Curtis and Bharucha, 2010; Ilie and Thompson, 2011).

Admittedly, perception tests have brought valuable insight into the acoustic relations between affective music and speech, but sometimes individual variation can significantly influence perceptual judgments which could lead to unreliable or even contradictory results (cf. Juslin and Sloboda, 2013). Therefore, this study aims to compare affective music and speech from a different perspective by using affective piano performance and speech production with a special focus on dynamics. This is because compared with the vast amount of perception studies, research using production experiments on affective music performance is on the rise only in the past 20 years, thanks to the advent of music technology that makes quantifying music performance easier than before. Particular interest has been given to piano/keyboard performance due to the availability of MIDI, 3D motion capture cameras, digital acoustic pianos (Palmer, 2006). However, to our knowledge, strictly controlled experiments that directly compare affective piano performance with speech production are rare. In this study we focus on dynamics in the comparison between the two domains with the inclusion of fingerings in piano performance and articulatory constraints in speech production. The reasons will be elaborated on in the following sections.

Dynamics of Piano Performance

In studies of affective piano performance, dynamics have received less attention than timing, although they are equally important (Repp, 1996; Gabrielsson, 2003). The reason is that unlike timing which can be easily measured by metronome and hence has been systematically examined in a scientific way for over a decade (Repp, 1992a,b, 1994a,b, 1995, among others), dynamics are more difficult to measure. This could be partly due to perceptual difficulty in precisely distinguishing different levels of dynamics (e.g., *forte* and *mezzoforte*) or technical challenges in filtering out unwanted acoustic artifacts (Repp, 1996).

Therefore in this study we decide to examine piano dynamics from a different perspective, i.e., at the kinematic level of dynamics which reflects “the varying forces of the pianist’s finger movements on the keyboard” (Repp, 1996, p. 642) by using a modified Moog PianoBar scanner (cf. McPherson, 2013). It is a portable scanner that can be rapidly attached to any acoustic piano keyboards. Using an optical reflectance sensing mechanism, the modified PianoBar scanner continuously detects key movements. Quantitatively, the scanner returns the values of continuous key positions (key displacement) and the time taken for fingers to reach each key position during one keystroke. As a result, multiple different dimensions of each key press, velocity and peak velocity (i.e., the maximum value in a continuous velocity trajectory) of key movement during each keystroke can be extracted from continuous key position data, following a similar approach to McPherson and Kim (2011). The multidimensions of key touch quantitatively returned by the scanner can provide an ideal platform for examining the

interaction between pianists’ expressive intention and their piano key touch (cf. McPherson and Kim, 2013).

Literature on mechanics of skilled motor movement (such as speech production and music performance) suggests that dynamics of motor movement are related not only to peak velocity but also to the movement amplitude, i.e., the peak velocity should be divided by the movement amplitude in order to compare dynamics of movement of different sizes (Nelson, 1983; Ostry et al., 1983; Ostry and Munhall, 1985). Therefore, in the context of piano performance, since each keystroke may correspond to different degrees of key displacement (i.e., different amplitudes of key movement), it is necessary to factor in key displacement at the point of peak velocity to yield the kinematic dynamics of each keystroke which reflects pianists’ finger force (Minetti et al., 2007). Similar approach can also be found in Kinoshita et al. (2007) where key displacement was taken as a factor in comparing finger force under the conditions of different types of key touch.

The examination of kinematic dynamics needs to take into account the role of fingerings. This is because in piano performance, alternative fingerings can be used for the same piece of music, which is unlike playing other instruments. Usually, different fingering strategies can reflect how pianists intend to interpret the structure, meaning and emotion of music in which dynamics play an important role (Neuhaus, 1973; Bamberger, 1976; Clarke et al., 1997). Parncutt et al. (1997) established a set of hypothetical rules of right-hand fingerings according to ergonomic difficulty such as the extent of hand spans, the involvement of weak fingers, and the coordinated playing on black and white keys. Of particular importance are hand spans and weak fingers. This is because the extent of hand spans can affect the degree of tension and physical effort of fingers (Parncutt et al., 1997). Weak fingers usually refer to the fourth and fifth fingers (Parncutt et al., 1997) which can constrain the flexibility of finger movement because of the hand’s anatomical structure: unlike the thumb and index fingers which are relatively independent, the middle, ring and little finger are closely linked to each other via the flexor digitorum profundus (FDP) tendons because they share a common muscle belly (Gunter, 1960). Moreover, the flexor digitorum superficialis (FDS) is especially responsible for the coupling between the fourth and fifth fingers (Baker et al., 1981; Austin et al., 1989). Nevertheless, whether weak fingers can significantly influence piano performance is still a matter of debate. As pointed out in Kochevitsky (1967), Neuhaus (1973), and Sandor (1981), weak fingers are not necessarily weak; instead, they are often strong enough to meet the demand of different levels of playing, especially octave playing.

Dynamics of Speech Production

With regard to speech, articulatory effort which reflects “force of articulation” (Malécot, 1955) is the counterpart of finger force in piano performance. Articulatory effort is essentially a neuromuscular phenomenon. Electrochemical reaction of nerve impulses triggers the activation of articulator muscles (Kirchner, 1998). Full contraction of articulator muscles occurs when agonist muscle activity outweighs the antagonist muscle

activity under the condition of repeated neuron firing (Clark and Yallop, 1990). Articulatory effort is therefore the sum action of the neuron firing of each articulator muscle (Kirchner, 1998). However, direct measurements of the neuron firing of each articulator muscle are clearly too intrusive and complex to perform. Therefore, indirect measurements have been put forth through examining phenomena related to articulatory gestures: articulator displacement (Lindblom, 1983), clear speech (Uchanski, 2008), fricative closures (Lavoie, 2001), trill articulation (Padgett, 2009), assimilation (Lindblom, 1983), all of which require great articulatory effort. Correspondingly, speech production models as in Lindblom and Sundberg (1971), Westbury and Keating (1986), Kirchner (1998), have been established in an attempt to quantify articulatory effort. However, the aforementioned measurements of articulatory gestures run the risk of not capturing articulatory phenomena large enough for statistically significant differences (Kaplan, 2010); in addition, the proposed models would oversimplify the reality of speech articulation which often involves much finer details than what the models can accommodate (Kaplan, 2010).

Hence, different alternatives are worth exploring. One such example is to use formant dynamics (i.e., trajectories and velocity) as an indicator of articulatory effort (cf. Cheng and Xu, 2013). Admittedly, one could argue formant dynamics may not be a reliable indicator of articulatory effort given the fact that there does not exist a one-to-one mapping between acoustics and articulation. Nevertheless, direct measurements on articulators as has been pointed out above do not capture the whole picture of articulatory movement either (cf. Cheng and Xu, 2013 for more examples and discussions). Acoustic signals, on the other hand, have been argued to provide reliable information for phonetic characteristics of segments and suprasegments with theoretical (Lindblom, 1990) and experimental evidence (Perkell et al., 2002). In addition, acoustic and articulatory measurements can produce similar dynamic patterns: the evidence is that the linear relations between F0/formant velocity and F0/formant movement amplitude (Xu and Wang, 2009; Cheng and Xu, 2013) in acoustics are similar to those in articulation (Kelso et al., 1985). Therefore, it is justifiable to use acoustic characteristics of formant dynamics to analyze articulatory effort.

In speech, formant patterns tend to be affected by articulatory constraints (e.g., articulatory pressure and distance) in different suprasegmental and segmental contexts (Erickson et al., 2004; Kong and Zeng, 2006). Tone languages such as Mandarin can be a typical platform for investigating articulatory pressure in different suprasegmental contexts: In Mandarin, there are five types of tones—High (tone 1), Rising (tone 2), Low (tone 3), Falling (tone 4), and Neutral (tone 5). Tone 2 + tone 2 and tone 4 + tone 4 create high articulatory pressure while tone 3 + tone 1 create weak articulatory pressure. The reason is that as reported in Xu and Wang (2009), successive rising tones (i.e., tone 2 + tone 2) or falling tones (tone 4 + tone 4) create much larger articulatory pressure than other tonal combinations because each involves two movements within one syllable. Successive static tones (tone 3 and tone 1), in contrast, have much smaller articulatory pressure because only a single movement is needed within each syllable. With regard to the segmental dimension,

diphthongs (i.e., two adjacent vowels) can be used because they are categorized into wide and narrow diphthongs according to their articulatory distance: wide diphthongs (e.g., /ai/, /au/, /ɔi/) have wider articulatory distance between the initial and final vowel and hence involve greater articulatory movement of speech organs. Narrow diphthongs (e.g., /ei/, /əu/) have narrower articulatory distance between the initial and final vowel and hence the articulatory movement is not as large as wide diphthongs.

Motivations for This Study

Theoretically, motion for a long time has been an important platform for investigating music, i.e., how physical motion is associated with sound patterns subsequently generated (Sundberg, 2000). Human voice is a direct reflection of such motion-to-sound mapping through physical coordination of articulatory gestures; meanwhile, performance of musical instruments is another way of mapping motion to sound through the use of tonguing, breathing, and fingering (Palmer et al., 2007, 2009). Therefore, similar to speech production, music performance can be conceptualized as a “sequence of articulatory movements resulting in a continuous acoustic wave” (Palmer et al., 2007, p. 119). In the context of piano performance, fingers can thus be perceived as “articulators” for pianists to articulate their interpretation of music. Indeed, experimental results on piano performance (Winges et al., 2013) show that speech production phenomenon such as coarticulation also exists in pianists’ finger movement during performance. This is not surprising given the fact that both speech production and piano performance are under neuromuscular control (Winges et al., 2013) and essentially both domains require skilled motor movements following similar physical mechanisms of dynamics (Grillner et al., 1982; Nelson, 1983; Ostry et al., 1983; Wings et al., 2013; van Vugt et al., 2014). In the context of motor movement, force is a crucial component contributing to the dynamics of physical movements (Stein, 1982). Therefore, it is reasonable to compare articulatory effort with force of other types of motor movements such as finger force (Gentil and Tournier, 1998; Ito et al., 2004; Loucks et al., 2010). As discussed in previous sections, the kinematic dynamics of keystroke reflect pianists’ finger force and the formant dynamics of speech reflect speakers’ articulatory effort. Since music performance and speech are two important platforms for humans to communicate emotion (Juslin and Laukka, 2003), plus the fact that these two domains are essentially skilled motor movements following similar physical mechanisms of dynamics as discussed above, it is therefore justifiable to compare music performance and speech production in the context of emotion using dynamics of motion (i.e., kinematic dynamics of keystroke and formant dynamics of speech production) as a measurement parameter. To our knowledge, such comparison is currently missing in literature and we believe it is worth bridging the gap.

In addition, one may wonder how piano fingerings (Section Dynamics of Piano Performance) and articulatory constraints (Section Dynamics of Speech Production) can relate to each other. Anatomically, articulation refers to motor movement caused by skeletal muscle contraction (Tortora, 2002). Hence

typical human motor movements such as speech production or music performance are effectively muscular articulation. There is no wonder, therefore, that pianists' fingers are always referred to as "articulators" expressing pianists' interpretation of music. Different fingerings involve different degrees of hand span and alternation between strong and weak fingers, which consequently lead to different degrees of finger muscular tension (Parncutt et al., 1997). Similarly in speech, different degrees of articulatory constraints are involved as discussed in Section Dynamics of Speech Production. Both finger muscular tension and speech articulatory pressure can be considered as physical constraints on motor movements such as piano performance and speech production (Nelson, 1983; Winges et al., 2013). Therefore it is the physical constraints triggered by different fingerings or articulatory constraints that relate the two domains to each other. Despite the importance of fingerings and articulatory constraints reviewed above, it is still unknown whether they interact with emotion in piano performance and speech production. This serves as another motivation for this study.

Four of the basic emotions (Ekman, 1992) are chosen: anger, fear, happiness, and sadness. One may wonder why a discrete model of emotion (Ekman, 1992; Panksepp, 1998) has been chosen rather than a dimensional approach such as Russell's circumplex model (1980). This is because firstly, so far no theoretical consensus has been reached as to which approach is better than the other for modeling emotion (for a recent summary of theoretical debates, see Zachar and Ellis, 2012). More importantly, the two approaches are not necessarily in conflict with each other as recent affective neuroscience studies (e.g., Panksepp and Watt, 2011) have suggested that the differences between the two may well be insignificant given the fact that both approaches share many common grounds in explaining cognitive functions of emotion. Since it is not the purpose of this study to test which model is better, a discrete model of affect is adopted. Among the "big six" emotions (Ekman, 1992), vocal disgust usually cannot be elicited satisfactorily under laboratory conditions (cf. Scherer, 2003); musical surprises can be very complicated often requiring sharp contrast in compositional structure (Huron, 2006) which is out of the scope of this study. Hence, only the remaining four of the "big six" emotions are chosen. The research questions to be addressed are.

Are dynamics of piano performance (i.e., finger force) similar to or different from dynamics of speech production (i.e., articulatory effort) under the condition of the four emotions? Do fingerings and articulatory constraints interact with emotion in their influence on the dynamics of piano performance and speech production respectively?

Experiments

Experiment 1: The Piano Experiment Stimuli

Two excerpts of music were composed for this study. According to the above reviews on fingerings, hand span and weak fingers should be the primary focus. Therefore, the two excerpts were composed corresponding to small and large hand span, respectively. Small hand span is where fingers are at their natural resting positions on the keyboard, i.e., without needing to extend far beyond the resting positions to reach the notes (Sandor, 1981). Large hand span is where fingers need to extend far beyond their resting positions, which usually involves stretching at least an octave (Parncutt et al., 1997). Meanwhile, each excerpt was to be played with strong finger combinations (the thumb, index, and middle fingers) and weak finger combinations (the ring and little fingers). In addition, given the fact that right and left hands tend to have different patterns in piano performance (Minetti et al., 2007), only the right hand is involved in this experiment to avoid theoretical and practical complexities. Hence altogether there are four levels of fingerings for this study: small-weak (SW), small-strong (SS), large-weak (LW), large-strong (LS). To avoid confounding effects, all excerpts have musically neutral structure, i.e., without having overtly emotional implications. Figures 1–4 demonstrate the fingering design.

Participants and Procedure

This experiment was approved by the Committee on Research Ethics at University College London. Eight professional pianists (four females, Mean = 26 years, $SD = 2.2$, all right-handed) from London were recruited to play the excerpts according to the fingerings provided on scores. They have been receiving professional piano training for an average of 20 years. They were instructed to play each of the excerpts with four emotions: anger, happiness, fear, and sadness. Each excerpt per emotion was repeatedly played three times in a quiet room. Admittedly, lacking ecological validity can be a problem with this method, i.e., it deviates from the reality of music making in that firstly, performances usually take place in concert halls; secondly, different emotions are often expressed by different pieces of music (cf. Juslin, 2001 for references therein). Nevertheless, real music making settings often cannot be scientifically controlled, i.e., it is impossible to filter out confounding factors coming from the acoustics of concert halls and audience. Moreover, it is hard to judge whether it is the way music is performed or the melody of music that leads the listeners to decide on the emotional categories if different pieces of music are used for

FIGURE 1 | The small-weak condition (SW): small hand span (i.e., fingers are at their natural resting positions) with only weak fingers, i.e., the ring (4) and little (5) fingers involved.



FIGURE 2 | The small-strong condition (SS): small hand span (i.e., fingers are at their natural resting positions) with only strong fingers, i.e., the thumb (1), index (2) and middle (3) fingers involved.



FIGURE 3 | The large-weak condition (LW): large hand span (i.e., fingers stretching across an octave) with only weak fingers, i.e., the ring (4) and little (5) fingers involved.



FIGURE 4 | The large-strong condition (LS): large hand span (i.e., fingers stretching across an octave) with only strong fingers, i.e., the thumb (1) and middle (3) fingers involved.

different emotions (Juslin, 2000, 2001). Therefore, conducting the experiment in a scientifically controlled way is still the better option if validity of the results is the priority.

As introduced in Section Dynamics of Piano Performance, a Moog PianoBar scanner was attached to the keyboard of a Bösendorfer grand piano. Finger force is reflected by keystroke dynamics which were calculated according to the formula: $dynamics = \frac{\text{peak velocity of each keystroke (Vp)}}{\text{maximum piano key displacement (d)}} (Vp/d)$ henceforth) because of the need to consider movement amplitude (i.e., displacement) in relation to peak velocity to reflect kinematic dynamics as reviewed above. More specifically, $Vp = \frac{\text{maximum piano key displacement (d)}}{\text{time taken to reach the maximum displacement (t)}}$, and so the ratio $Vp/d = \frac{d}{t} \times \frac{1}{d} = \frac{1}{t}$. The unit of displacement is mm and that of time is sec. The data were obtained by an external computer attached to one end of the PianoBar. A Matlab script was written for computing dynamics according to the formula.

There were altogether 8 (pianists) * 4 (emotions) * 4 (fingerings) * 3 (repetitions) = 384 episodes performed by the pianists. A follow-up perceptual validation test was carried out: sixteen professional musicians (10 females, Mean = 28 years, $SD = 1.5$) were asked to rate each emotion * fingering episode on a 1–5 scale. 1 represents not at all angry/fearful/happy/sad while 5 represents very angry/fearful/happy/sad. The top 8 ranked episodes (out of 24) for each emotion * fingering were selected. The mean score for each emotion * fingering was 4.03.

TABLE 1 | Mean Vp/d of the four levels of emotion (A, anger; F, fear; H, happiness; S, sadness) and the four levels of fingerings (SW, small-weak; SS, small-strong; LW, large-weak; LS, large-strong).

	Anger	Fear	Happiness	Sadness
Mean Vp/d	25.2	17.3	22.9	5.8
Standard deviation	1.8	4.9	2.1	1.2
	SW	SS	LW	LS
Mean Vp/d	13.7	17.8	19.4	20.5
Standard deviation	2	1.8	2.6	1.5

Results of the Piano Experiment

A Two-Way repeated measures ANOVA was performed to examine the effect of emotion (four levels: anger, fear, happiness, and sadness) and fingerings (four levels: small-weak, small-strong, large-weak, large-strong). The results (Table 1) demonstrate that both factors play significant roles in finger force reflected by Vp/d. The interaction between the two factors is also significant.

The means of keystroke dynamics (Vp/d) for each condition are displayed in Table 2 and *Post-hoc* Tukey HSD tests (Table 3) reveal more detailed patterns: anger and happiness have significantly higher dynamics than fear and sadness. The differences between anger and happiness are non-significant.

TABLE 2 | Results of the Two-Way repeated-measures ANOVA of emotion and fingerings on keystroke dynamics (as reflected by Vp/d).

	F	df	p	η_p^2
Emotion	8.26	3.21	<0.001	0.31
Fingerings	4.05	3.21	<0.05	0.16
Emotion * fingerings	2.17	9.63	<0.05	0.13

TABLE 3 | Results of Post-hoc Tukey tests on means of the four levels of emotion (A, anger; F, fear; H, happiness; S, sadness) and the four levels of fingerings (SW, small-weak; SS, small-strong; LW, large-weak; LS, large-strong).

ρ	A vs. F	A vs. H	A vs. S	F vs. H	F vs. S	H vs. S
	<0.05	>0.05	<0.001	<0.05	<0.05	<0.001
ρ	SW vs. LS	SW vs. LW	SW vs. SS	LS vs. LW	LS vs. SS	LW vs. SS
	<0.01	<0.01	<0.05	>0.05	>0.05	>0.05

Fear has significantly lower dynamics than anger and happiness but it is still significantly higher in dynamics than sadness. With regard to the factor of fingerings, the Tukey tests demonstrate that weak fingers in large hand span (the LW condition) do not produce significantly different dynamics from strong fingers in large hand span (the LS condition). However, under the condition of small hand span, weak fingers tend to produce significantly lower dynamics than strong fingers.

In terms of the interaction between emotion and fingerings, **Figure 5** shows that the most obvious interaction is between fear, large-strong (LS), large-weak (LW), and small-strong (SS) fingering conditions. For all of the aforementioned fingerings, fear has significantly higher ($p < 0.05$) dynamics than sadness according to a series of *post-hoc* Tukey tests, although it is still significantly lower ($p < 0.05$) than anger and happiness. Between the LS, LW, and SS conditions in fear, the differences are non-significant. For anger, happiness and sadness, large hand span generates higher dynamics than small hand span, but the differences are non-significant, i.e., regardless of whether the hand span is large or small, the dynamics are on average always high for anger and happiness while for sadness they are always low. Therefore, the contrast in dynamics between different fingerings is evident under the condition of fear only.

Experiment 2: The Speech Experiment Stimuli

The stimuli consist of six sentences divided into two sets (Tables 4, 5), with tones and vowels being the two variables. The purpose is to use the two variables to test the respective impact of suprasegmental and segmental constraints on formant dynamics. According to the reviews in Section Dynamics of Speech Production, tone 2 + tone 2 and tone 4 + tone 4 are used to create high articulatory pressure. Tone 3 + tone 1 is used to create low articulatory pressure. Meanwhile, a wide diphthong /au/ is used for long segmental distance and a narrow diphthong /əu/ is used for short segmental distance. *Cuilaoyou* and *Cuilouyou* are compound words denoting a person's name.

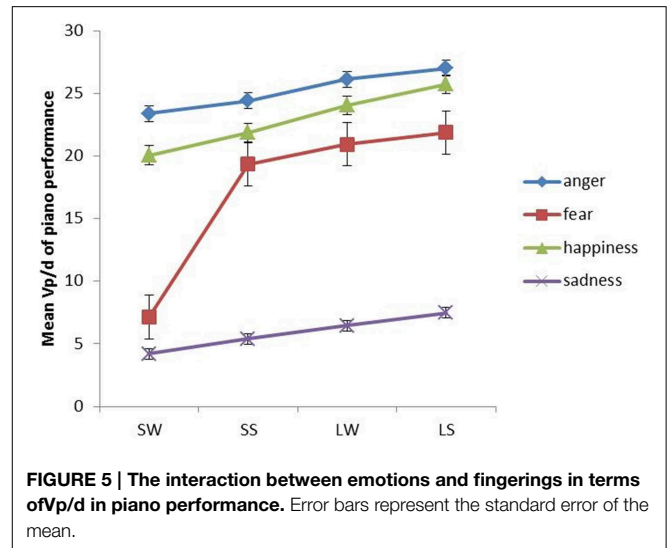


FIGURE 5 | The interaction between emotions and fingerings in terms of Vp/d in piano performance. Error bars represent the standard error of the mean.

TABLE 4 | The first set of the stimuli in which the numbers of the syllables represent the five lexical tones in Mandarin: 1 for H (High tone), 2 for R (Rising tone), 3 for L (Low tone), 4 for F (Falling tone), and 5 for N (Neutral tone).

	lao2[<i>lau</i>]	yao2[<i>jau</i>]				
	work	distant				
cui1	lao3[<i>lau</i>]	yao1[<i>jau</i>]	nian4	shu1	qu4	le5
surname	old	waist	read	book	aspect	particle
	lao4[<i>lau</i>]	yao4[<i>jau</i>]				
	flood	medicine				

IPA transcriptions for the key words *laoyao* are provided in brackets. Translation: *cuilaoyou* has gone to read a book.

TABLE 5 | The second set of the stimuli in which the numbers of the syllables represent the five lexical tones in Mandarin: 1 for H (High tone), 2 for R (Rising tone), 3 for L (Low tone), 4 for F (Falling tone), and 5 for N (Neutral tone).

	lou2[<i>ləu</i>]	you2 [<i>jəu</i>]				
	building	oil				
cui1	lou3[<i>ləu</i>]	you1[<i>jəu</i>]	nian4	shu1	qu4	le5
surname	hug	good	read	book	aspect	particle
	lou4[<i>ləu</i>]	you4[<i>jəu</i>]				
	drip	right				

IPA transcriptions for the key words *louyou* are provided in brackets. Translation: *cuilouyou* has gone to read a book.

Measurement of Formant Dynamics

As reviewed in Section Dynamics of Speech Production, formant dynamics are an important factor reflecting the articulatory effort of speech production. Formant peak velocity, i.e., “the highest absolute value in the continuous velocity profile of the (formant) movement” (Cheng and Xu, 2013, p. 4488), and the displacement/amplitude of the formant movements are particularly related to articulatory effort [cf. Cheng and Xu (in press) for further discussion]. The peak velocity is measured in the following way (Xu and Wang, 2009, p. 506):

“Positive and negative extrema in the velocity curve correspond to the rising and falling ramps of each unidirectional pitch (formant) movement. A velocity curve was computed by taking the first derivative of an F0 (formant) curve after it has been smoothed by low-pass filtering it at 20 Hz with the Smooth command in Praat. Following Hertrich and Ackermann (1997), the velocity curve itself was not smoothed so as not to reduce the magnitude of peak velocity.”

Figures 6, 7 show the measurement points taken from F1 and F2 formant contours. This allows the calculation of the ratio of formant peak velocity (V_p) to maximum formant displacement (d), henceforth V_p/d . It reflects articulatory effort/vocal vigorousness (Cheng and Xu, 2013, in press). Similar to the piano experiment, $V_p = \frac{\text{maximum formant displacement } (d)}{\text{time taken to reach the maximum displacement } (t)}$, and so the ratio $V_p/d = \frac{d}{t} \times \frac{1}{d} = \frac{1}{t}$. The unit of displacement is Hz and the unit of time is sec.

Table 6 lists the values extracted from the measurement points for the calculation of V_p/d for F1 and F2.

Subjects and Procedure

Ten native Mandarin speakers without speech or hearing problems were recruited as subjects (5 females; Mean = 27 years, $SD = 2.5$) via the University College London Psychology

Pool. The recording session for each participant lasted for around half an hour. This experiment was approved by the Committee on Research Ethics at University College London. Voice portrayal/simulation method was used to induce emotions, i.e., the participants were asked to imagine themselves in emotion-triggering scenarios when recording the sentences.

TABLE 6 | Values taken from the measurement points a, b, c for the calculation of V_p/d .

minF1a: F1 minimum in a	minF2a: F2 minimum in b
maxF1a: F1 maximum in b	maxF2a: F2 maximum in a
minF1b: F1 minimum in b-c	minF2b: F2 minimum in c
maxF1b: F1 maximum in c	maxF2b: F2 maximum in b-c
D1a: maxF1a – minF1a [F1 rising displacement]	D2a: maxF2a – minF2a [F2 falling displacement]
D1b: maxF1a – minF1b [F1 falling displacement]	D2b: maxF2b – minF2a [F2 rising displacement]
D1c: maxF1b – minF1b [F1 rising displacement]	D2c: maxF2b – minF2b [F2 falling displacement]
V1a: F1 peak rising velocity, in a-b	V2a: F2 peak falling velocity, in a-b
V1b: F1 peak falling velocity, in a-b	V2b: F2 peak rising velocity, in a-b
V1c: F1 peak rising velocity, in b-c	V2c: F2 peak falling velocity, in b-c

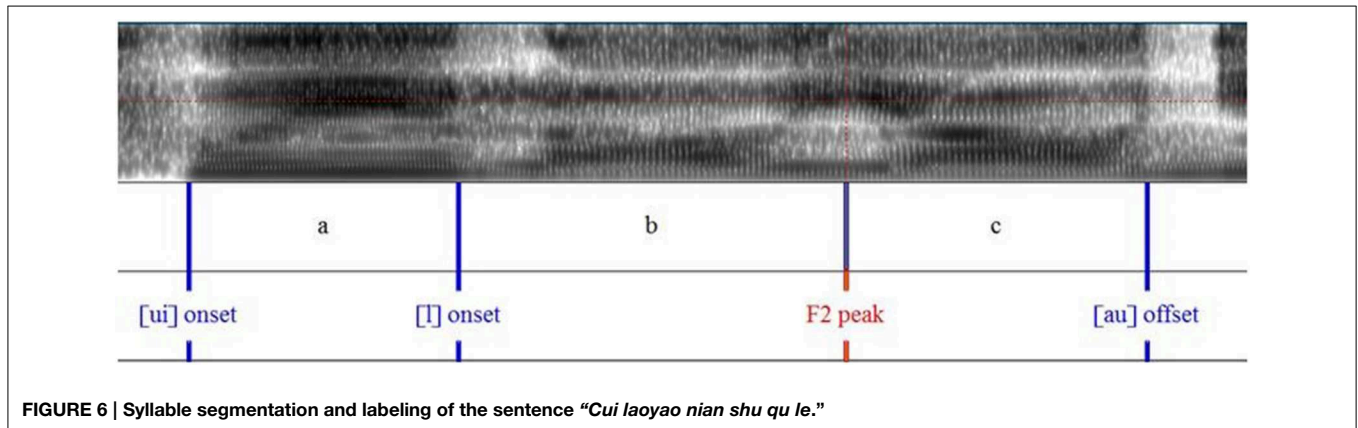


FIGURE 6 | Syllable segmentation and labeling of the sentence “Cui laoyao nian shu qu le.”

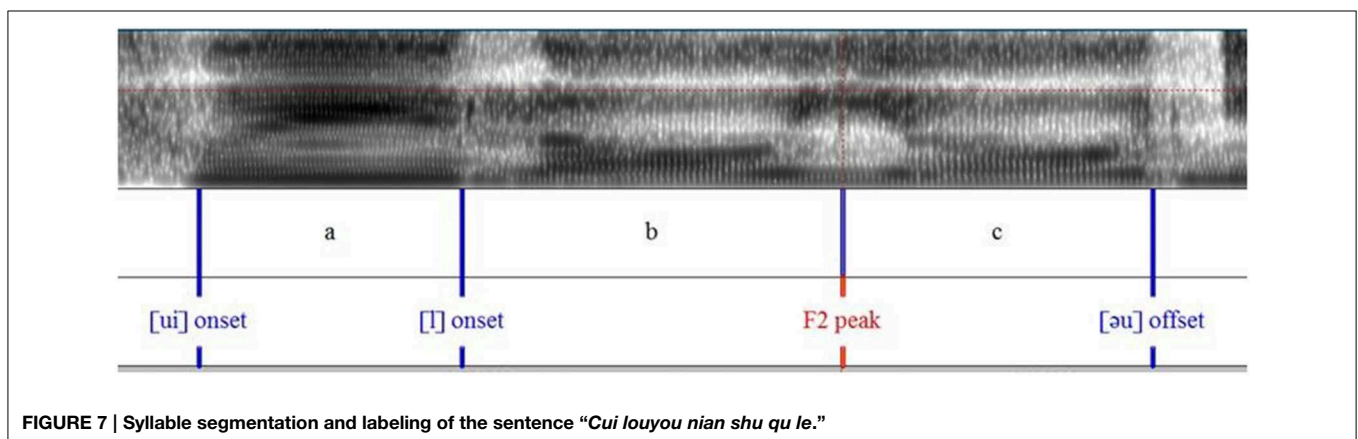


FIGURE 7 | Syllable segmentation and labeling of the sentence “Cui louyou nian shu qu le.”

This is because compared to other emotional speech induction methods (e.g., natural vocal expression), this method is more effective in obtaining relatively genuine emotional speech when experimental control is a key concern. Support for this method comes from the fact that natural emotional expression is often inherently involved with unintended portrayal and self-representation (Scherer, 2003). The recording was conducted in a sound-controlled booth. Participants were asked to record each sentence 3 times in four emotions: anger, fear, happiness, and sadness, resulting in 10 (speakers) * 4 (emotions) * 3 (tones) * 2 (segments) * 3 (repetitions) = 720 tokens.

Similar to the first experiment, a follow-up perception validation test was conducted: twenty native speakers of Mandarin (11 females, Mean = 23 years, SD = 2.6) were asked to rate each emotion * tone * segment token on a 1–5 scale. 1 represents not at all angry/fearful/happy/sad while 5 represents very angry/fearful/happy/sad. The top eight ranked tokens (out of 30) for each emotion * tone * segment were selected. The mean score for each emotion * tone * segment was 4.16. ProsodyPro and FormantPro scripts (Xu, 2014) running under Praat (Boersma and Weenink, 2013) was used for data analyses.

Results

The mean Vp/d of all measurement points are represented in Table 7. A Three-Way repeated measures ANOVA shows that among the three factors (emotion, tone, and segments), emotion is the only factor exerting a significant impact on the value of Vp/d. The interaction between emotion, tone and segments is non-significant. However, the interactions between emotion and tone and that between emotion and segments are significant (Table 8).

Post-hoc Tukey tests show more details: sadness has significantly ($p < 0.05$) the lowest Vp/d value compared with

TABLE 7 | Mean Vp/d of the four levels of emotion (anger; fear; happiness; sadness) and the four levels of articulatory constraints (SS, small segmental distance /əu/; ST, small tonal pressure T3 + T1; LS, large segmental distance /au/; LT, large tonal pressure T2 + T2/T4 + T4).

	Anger	Fear	Happiness	Sadness
Mean Vp/d	37.3	41.5	43.6	24
Standard deviation	6.2	5.5	5.6	4.9
	SS	ST	LS	LT
Mean Vp/d	34	33.6	38.5	39.8
Standard deviation	5.9	4.8	5.1	5

TABLE 8 | Results of the Three-Way repeated-measures ANOVA on articulation dynamics (as reflected by Vp/d).

	F	Df	P	η^2_p
Emotion	5.22	3,21	<0.01	0.25
Emotion * tone	2.39	6,42	<0.05	0.11
Emotion * segments	3.08	3,21	<0.05	0.12

the other three emotions. Happiness has the highest dynamics followed by fear and anger, but the differences between each other are non-significant.

The interaction between emotion and tonal pressure is significant. As shown in Figure 8, the Vp/d of all emotions is higher in tonal combinations of large articulatory constraints (i.e., T2 + T2 and T4 + T4) than the Vp/d in those of small articulatory constraints (T3 + T1). This is the most obvious in the case of anger where T2 + T2 and T4 + T4 make the Vp/d of anger become closer to that of fear and happiness. Post-hoc Tukey tests show that the difference between anger and fear plus that between anger and happiness are non-significant under the T2 + T2 and T4 + T4 conditions. In contrast, under the T3 + T1 condition, the differences are significant (both $ps < 0.05$). In addition, fear, happiness and sadness do not differ significantly between the two tonal conditions. Therefore, anger is more affected by tonal variation than the other three emotions.

Emotion also interacts significantly with segmental distance. Figure 8 shows Vp/d is overall higher in the wide diphthong condition than in the narrow condition. The interaction is the most obvious in anger because it is almost as high as fear and happiness with regard to Vp/d in the wide diphthong condition. Post-hoc Tukey tests show the difference between anger and fear plus that between anger and happiness are non-significant in the wide diphthong condition. The differences are significant (both $ps < 0.05$), however, under the narrow condition. Moreover, fear, happiness and sadness do not differ significantly between the two segmental distance conditions. Therefore, similar to above, anger is more influenced by segmental distance variation than the other three emotions.

Comparisons between the Results of the Piano and Speech Experiment

To directly compare the results of the piano and speech experiments, a MANOVA test was conducted: the within-subjects independent variables are emotion (four levels: anger, fear, happiness, and sadness) and physical constraint (two levels: large hand span/articulatory constraints and small

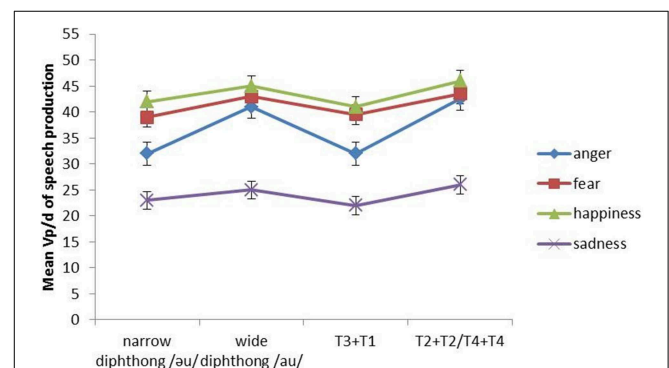


FIGURE 8 | The mean Vp/d of narrow diphthong, wide diphthong, T3 + T1 and T2 + T2/T4 + T4 in the four types of emotional speech (anger, fear, happiness, and sadness). Error bars represent the standard error of the mean.

hand span/articulatory constraints) while the between-subjects independent variable is group (two levels: pianists and speakers). The dependent variables are Vp/d (speakers) and Vp/d (pianists). Using Pillai's trace, there is a significant difference between pianists and speakers [$F_{(8, 7)} = 13.78, p < 0.01$]. The following univariate ANOVAs show that the group differences between pianists and speakers are significant across most conditions: anger-large [$F_{(1, 14)} = 14.92, p < 0.01$], anger-small [$F_{(1, 14)} = 16.23, p < 0.01$], happiness-small [$F_{(1, 14)} = 15.61, p < 0.01$], fear-large [$F_{(1, 14)} = 14.95, p < 0.01$], fear-small [$F_{(1, 14)} = 18.09, p < 0.01$], sadness-large [$F_{(1, 14)} = 15.93, p < 0.01$]. In the happiness-large and sadness-small conditions, the group difference is non-significant (speech production still has higher Vp/d than that of piano performance). The results suggest on the whole, piano performance has significantly different (i.e., lower) dynamics than speech production.

Discussion

Similarities between Affective Piano Performance and Speech Production

The results show that firstly, anger in piano performance generates the highest dynamics irrespective of fingerings; in speech production, it is also relatively high in dynamics although it interacts with articulatory constraints (more discussions on the interaction are offered in the following section). This is in line with previous reports that anger in music performance and speech production is generally linked to fast tempo, high intensity and great energy (cf. Juslin and Sloboda, 2013). Physiologically, the high dynamics of anger can be associated with high levels of cardiovascular activities such as high heart rate (Rainville et al., 2006), fast/deep breathing (Boiten et al., 1994), increases in diastolic pressure and activated baroreceptor mechanisms (Schwartz et al., 1981). Evolutionarily, anger originates from natural and sexual selection pressure on animals (Darwin, 1872): anger induces the inclination to fight or attack whatever that threatens survival and well-being. As a result, anger is proposed to be associated with large body size projection (Morton, 1977; Xu et al., 2013a,b) to scare off enemies. Hence anger should be linked to high dynamics which can be reflected by high physical or vocal effort to show great strength and energy (Xu et al., 2013a). The results of this study support this prediction by demonstrating that greater finger force and articulatory effort are generated respectively in piano performance and speech production in the context of anger.

Secondly, happiness triggers the highest dynamics for speech production and second highest dynamics for piano performance, irrespective of fingerings or articulatory constraints. The results are in line with previous reports that in music performance, happiness is always associated with faster tempo and higher intensity (Gabrielsson, 1995; Zanon and De Poli, 2003a,b; Widmer and Goebel, 2004); happy speech is reported to have high values in many acoustic dimensions such as pitch, pitch range, intensity, high frequency energy (Scherer, 2003; Ververidis and Kotropoulos, 2006), speech rate, and formant shift (Xu et al., 2013a). Similar to anger, the physiological reason for high

dynamics of happiness is often linked to increases in heart rate, blood pressure, breathing pattern (Boiten et al., 1994; Rainville et al., 2006), all of which can contribute to greater physical or vocal force in music performance or speech production. From an evolutionary perspective, happiness can be a useful strategy for attracting mates (Darwin, 1872). Therefore, it is beneficial for sound signalers to produce highly vigorous (i.e., dynamic) sounds so as to be audible to potential mates (Xu et al., 2013a). Hence, the results are also consistent with the evolutionary account.

Thirdly, fear in both piano performance and speech production produces significantly higher dynamics than sadness; particularly in speech production fear does not differ significantly from anger/happiness. This might seem somewhat unexpected given that fear in music performance is generally associated with soft playing similar to sadness (cf. Juslin and Sloboda, 2013). In terms of speech production, however, fear has already been found to show high dynamics (Xu et al., 2013a), which is consistent with the view that evolutionarily, fear can be a defensive emotion (LeDoux, 1996), evidenced from animal alarm calls as a useful antipredator defensive strategy across many species for the sake of group survival (Caro, 2005). To serve this purpose, alarm calls should be reasonably high in dynamics (i.e., vigorousness). Similarly, musical excerpts of fear could also be highly dynamic, analogous to human fearful speech or animal alarm calls.

Fourthly, sadness always generates the lowest dynamics for both piano and speech performance regardless of fingerings or articulatory constraints. This finding is in line with previous research: sad music and speech are low in intensity, F0, F0 range and duration (Juslin and Laukka, 2003; Laukka et al., 2005; Patel, 2008). This is mainly because sadness is located at the opposite end of happiness in terms of valence and arousal: it is a lowly aroused negative emotion because of its association with reduced physiological energy and arousal level, sometimes leading to affective pathology such as depression or anhedonia (cf. Huron, 2011). Evolutionarily, such low dynamics of sadness indicate a tendency for the sound signaller to beg for sympathy (Xu et al., 2013a). Hence usually low motor effort is involved in expression of sadness either through music or speech. It is worth mentioning sad speech can be split into two categories: depressed sadness and mourning sadness (Scherer, 1979), the former being characterized by low vocal energy while the latter by high vocal energy. In this study, it was the depressed sadness that was used and hence the resulting formant dynamics are low, reflecting decreased articulatory effort due to the sluggishness of articulatory muscles in sad speech (Kienast and Sendlmeier, 2000).

Differences between Affective Piano Performance and Speech Production

The results also show significant differences between the two domains. The most notable difference is that speech production on the whole has higher dynamics than piano performance across almost all conditions. This is consistent with previous studies on comparisons between speech articulatory movements and limb movements (Gentil and Tournier, 1998; Ito et al., 2004; Loucks et al., 2010). Although those studies did not investigate movements in the context of affective piano performance

or speech production, the general biophysical mechanisms of fingers and speech articulators apply to this study. More specifically, it was found that compared with fingers or arms, speech articulators in general produce faster velocity (Ito et al., 2004; Loucks et al., 2010) and greater force (Gentil and Tournier, 1998; Loucks et al., 2010). The reasons probably lie in the biomechanical differences between speech articulators and fingers: compared with speech articulators, fingers are associated with more intervening factors (e.g., long tendons, joints and muscle mass between muscle fibers and skeletal joints) that prevent finger muscles from contracting as fast as speech articulatory muscles (Gentil and Tournier, 1998). It has also been reported that oral-facial muscles are associated with fast-twitch fibers and motor protein such as myosin which enable fast acceleration and rapid speech in order to meet different levels of speech demand (Williams and Warwick, 1980; Burke, 1981). Therefore, in this study the dynamics of affective speech production (as reflected by articulatory effort) and piano performance (as reflected by finger force) are different from each other due to the biomechanical distinctions.

In addition, the results also demonstrate that the interaction between emotion and physical constraints in piano performance is different from that in speech production. In piano performance (Figure 5), only fear interacts with physical constraints (i.e., fingerings); in speech production (Figure 8), only anger interacts with physical constraints (i.e., articulatory constraints). The reasons could be attributed to the differences in the extent of acoustic stability of music performance and speech production in different emotions.

Firstly, in music performance, anger, happiness, and sadness are associated with relatively consistent acoustic patterns (Juslin and Sloboda, 2013), i.e., anger and happiness are always fast and loud to convey high energy and arousal while sadness is always slow and quiet to convey low energy and arousal. Fear, in contrast, is linked to highly variable acoustic patterns especially in terms of tempo and intensity (Juslin and Madison, 1999; Madison, 2000; Juslin and Sloboda, 2013; Bernays and Traube, 2014) so as to convey the unstable psychological state under the influence of fear, e.g., scattered notes with pauses between musical phrases and sharp contrasts between intensity are often used to express fear (Madison, 2000). This could further imply there may not be a consistent pattern of finger force under the condition of fear. Hence, other factors such as fingerings are highly likely to interact with fear to generate different kinematic dynamics in piano performance.

On the other hand, fearful speech shown in this study always has high formant dynamics regardless of articulatory constraints. This is likely associated with duration. Upon close examination, the duration of fear (mean = 555.6 ms) is similarly short to that of happiness (mean = 546.6 ms) which has the highest dynamics, with non-significant differences between the two. Moreover, fear is significantly ($p < 0.05$) shorter than anger (mean = 601.1 ms) and sadness (mean = 638.2 ms). Similar findings have been reported that fear is often produced with fast speech rate that is likely to trigger vowel undershoot (i.e., an articulatory phenomenon where the canonical phonetic forms of speech sounds fail to be reached because of the articulatory

impact of surrounding segments, Lindblom, 1963) and segmental reduction (Paeschke et al., 1999; Kienast and Sendlmeier, 2000). Shorter duration is highly likely to trigger great articulatory effort according to the report of studies on articulatory movement (Munhall et al., 1985; Ostry and Munhall, 1985; Edwards et al., 1991; Adams et al., 1993; Perkell et al., 2002). Therefore, the relatively stable acoustic pattern (i.e., duration) of fearful speech could make it less likely to interact with other factors such as articulatory constraints.

Secondly, this study shows that only angry speech significantly interacts with articulatory constraints: the formant dynamics are significantly higher in large articulatory constraints than those in small articulatory constraints. Again this can be linked to duration. A closer look at the data reveals that the duration of angry speech is significantly ($p < 0.05$) shorter under the condition of large articulatory constraints than the condition of small articulatory constraints. It has been reported (Cheng and Xu, 2013) that when time is short for the articulatory execution of segments with large articulatory constraints, muscles have to contract faster (i.e., with stronger articulatory effort) than when small articulatory constraints are involved in order to reach the tonal and segmental targets. This is reflected in the high formant dynamics under the condition of large articulatory constraints. In addition, the result is also consistent with the finding that anger is often more variable in duration compared with the other three emotions (happiness, fear and sadness): it can be slow because of the need to be precise and clear in articulation (Paeschke et al., 1999; Kienast and Sendlmeier, 2000) so as to project big body size to threaten away enemies (Xu et al., 2013a,b); it can also be fast in speech rate (Scherer, 2003) especially in female speakers to reflect the highly aroused and variable psychological state under the influence of anger. Hence, it is the relatively high variability in duration that makes angry speech more prone to interact with external factors such as articulatory constraints.

Conclusion

This study compares the dynamics of piano performance (i.e., finger force) and those of speech production (i.e., articulatory effort) in four emotions: anger, happiness, fear and sadness. The results show that firstly, in both piano performance and speech production, anger and happiness generally have high dynamics while sadness has the lowest dynamics. The findings echo the theoretic argument that affective music shares a “common code” with affective speech (Juslin and Laukka, 2003). Secondly, the interaction between fear and fingerings in piano performance and the interaction between anger and articulatory constraints in speech production suggest that the more variable an emotion is in acoustic features, the more likely it is to interact in production with external factors such as fingerings or articulatory constraints in terms of dynamics. In addition, the results suggest that affective speech production on the whole has higher dynamics than affective piano performance, which may be due to the biomechanical differences between speech articulators and fingers.

Therefore, this is the first study to quantitatively demonstrate the importance of considering motor mechanisms such as

dynamics (i.e., finger force and articulatory effort) together with physical constraints (i.e., fingerings and articulatory constraints) in examining the similarities and differences between affective music performance and speech production. Limitations also exist: The emotion induction method of the piano and speech experiment still needs improvement due to lack of authenticity under the laboratory condition. Moreover, more fingering strategies especially those involving black keys and more articulatory variations in speech such as monophthongs vs. diphthongs could be added to research designs for more comprehensive findings. In addition, more categories of emotion such as disgust and sarcasm could be included to make the picture more complete. In a nutshell, focusing on the motor mechanisms

of affective music performance and speech production could further enhance the connection between music and speech as two fundamental capacities for humans to communicate emotion.

Acknowledgments

We would like to thank Dr. Andrew McPherson for his generous help with piano data acquisition using PianoBar and Dr. Jyrki Tuomainen for helpful discussion. We would also like to thank the two reviewers and the associate editor for insightful comments. Our gratitude also goes to pianists from London and Chinese students at the University College London for their enthusiastic participation in the experiments.

References

- Adams, S. G., Weismer, G., and Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *J. Speech Hear. Res.* 36, 41–54.
- Austin, G. J., Leslie, B. M., and Ruby, L. K. (1989). Variations of the flexor digitorum superficialis of the small finger. *J. Hand Surg.* 14A, 262–267.
- Baker, D. S., Gaul, J. S. Jr., Williams, V. K., and Graves, M. (1981). The little finger superficialis-clinical investigation of its anatomic and functional shortcomings. *J. Hand Surg.* 6, 374–378.
- Bamberger, J. (1976). “The musical significance of Beethoven’s fingerings in the piano sonatas,” in *Music Forum*, Vol. 4, ed F. Salzer (New York, NY: Columbia University Press), 237–280.
- Bernays, M., and Traube, C. (2014). Investigating pianists’ individuality in the performance of five timbral nuances through patterns of articulation, touch, dynamics, and pedaling. *Front. Psychol.* 5:157. doi: 10.3389/fpsyg.2014.00157
- Boersma, P., and Weenink, D. (2013). *Praat: Doing Phonetics by Computer [Computer Program]. Version 5.3.59*. Retrieved from: <http://www.praat.org/>
- Boiten, F. A., Frijda, N. H., and Wientjes, C. J. E. (1994). Emotions and respiratory patterns: review and critical analysis. *Int. J. Psychophysiol.* 17, 103–128.
- Buck, R. (1984). *The Communication of Emotion*. New York, NY: Guilford Press.
- Burke, R. E. (1981). “Motor units: anatomy, physiology and functional organization”, in *Handbook of Physiology. Section 1: The Nervous System. Vol. II. Motor Control, Part 1*, ed V. B. Brooks (Washington, DC: American Physiological Society), 345–422.
- Carlo, N. S., and Guaitella, I. (2004). Facial expressions of emotion in speech and singing. *Semiotica* 149, 37–55. doi: 10.1515/semi.2004.036
- Caro, T. (2005). *Antipredator Defenses in Birds and Mammals*. Chicago, IL: University of Chicago Press.
- Cheng, C., and Xu, Y. (2013). Articulatory limit and extreme segmental reduction in Taiwan Mandarin. *J. Acoust. Soc. Am.* 134, 4481–4495. doi: 10.1121/1.4824930
- Cheng, C., and Xu, Y. (in press). Mechanism of disyllabic tonal reduction in Taiwan Mandarin. *Lang. Speech*. doi: 10.1177/0023830914543286
- Clark, J., and Yallop, C. (1990). *An Introduction to Phonetics and Phonology*. Oxford: Blackwell.
- Clarke, E. F., Parncutt, R., Raekallio, M., and Sloboda, J. A. (1997). Talking fingerings: an interview study of pianists’ views on fingering. *Music Sci.* 1, 87–107.
- Cross, I., Fitch, W. T., Aboitiz, F., Iriki, A., Jarvis, E. D., Lewis, J., et al. (2013). “Culture and evolution,” in *Language, Music, and the Brain: A mysterious relationship*, ed M. A. Arbib (Cambridge, MA: MIT Press), 541–562.
- Curtis, M. E., and Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion* 10, 335–348. doi: 10.1037/a0017928
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.
- Edwards, J. R., Beckman, M. E., and Fletcher, J. (1991). The articulatory kinematics of final lengthening. *J. Acoust. Soc. Am.* 89, 369–382.
- Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6, 169–200.
- Erickson, D., Iwata, R., Endo, L., and Fujino, A. (2004). “Effect of tone height on jaw and tone articulation in Mandarin Chinese,” in *Proceedings of the International Symposium on Tonal Aspects of Languages* (Beijing), 53–56.
- Fonagy, I., and Magdics, K. (1963). Emotional patterns in intonation and music. *Z. Phon.* 16, 293–326.
- Gabrielsson, A. (1995). “Expressive intention and performance,” in *Music and the Mind Machine*, ed R. Steinberg (New York, NY: Springer), 35–47.
- Gabrielsson, A. (2003). Music performance research at the millennium. *Psychol. Music* 31, 221–272. doi: 10.1177/03057356030313002
- Gentil, M., and Tournier, C. L. (1998). Differences in fine control of forces generated by the tongue, lips and fingers in humans. *Arch. Oral Biol.* 43, 517–523.
- Grillner, S., Lindblom, B., Lubker, J., and Persson, A. (1982). *Speech Motor Control*. New York, NY: Raven Press.
- Gunter, G. S. (1960). Traumatic avulsion of the insertion of flexor digitorum profundus. *Aust. NZ. J. Surg.* 30, 1–9.
- Hertrich, I., and Ackermann, H. (1997). Articulatory control of phonological vowel length contrasts: kinematic analysis of labial gestures. *J. Acoust. Soc. Am.* 102, 523–536. doi: 10.1121/1.419725
- Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
- Huron, D. (2011). Why is sad music pleasurable? A possible role for prolactin. *Music Sci.* 15, 146–158. doi: 10.1177/1029864911401171
- Ilie, G., and Thompson, W. F., (2011). Experiential and cognitive changes following seven minutes exposure to music and speech. *Music Percept.* 28, 247–264. doi: 10.1525/MP.2011.28.3.247
- Ito, T., Murano, E. Z., and Gomi, H. (2004). Fast force-generation dynamics of human articular muscles. *J. Appl. Physiol.* 96, 2318–2324. doi: 10.1152/jappphysiol.01048.2003
- Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. *J. Exp. Psychol. Hum. Percept. Perform* 26, 1797–1812. doi: 10.1037/0096-1523.26.6.1797
- Juslin, P. N. (2001). “Communicating emotion in music performance: a review and a theoretical framework,” in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 309–337.
- Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770
- Juslin, P. N., and Madison, G. (1999). The role of timing patterns in the decoding of emotional expressions in music performances. *Music Percept.* 17, 197–221.
- Juslin, P. N., and Sloboda, J. A. (2013). “Music and emotion,” in *The Psychology of Music, 3rd Edn.*, ed D. Deutsch (Amsterdam: Elsevier), 583–645.
- Kaplan, A. (2010). *Phonology Shaped by Phonetics: The Case of Intervocalic Lenition*. Ph.D. dissertation. Santa Cruz, CA: University of California.
- Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., and Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modelling. *J. Acoust. Soc. Am.* 77, 266–280.

- Kienast, M., and Sendlmeier, W. F. (2000). "Acoustical analysis of spectral and temporal changes in emotional speech," in *Proceedings of the ISCA ITRW on Speech and Emotion* (Belfast), 92–97.
- Kinoshita, H., Furuya, S., Aoki, T., and Altenmüller, E. (2007). Loudness control in pianists as exemplified in keystroke force measurements on different touches. *J. Acoust. Soc. Am.* 121, 2959–2969. doi: 10.1121/1.2717493
- Kirchner, R. (1998). *An Effort-Based Approach to Consonant Lenition*. Ph.D. dissertation. UCLA, Los Angeles.
- Kochevitsky, G. (1967). *The Art of Piano Playing: A Scientific Approach*. Princeton, NJ: Summy-Birchard Music.
- Kong, Y. Y., and Zeng, F. G. (2006). Temporal and spectral cues in Mandarin tone recognition. *J. Acoust. Soc. Am.* 120, 2830–2840. doi: 10.1121/1.2346009
- Laukka, P., Juslin, P. N., and Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cogn. Emot.* 19, 633–653. doi: 10.1080/02699930441000445
- Lavoie, L. M. (2001). *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. New York, NY: Garland.
- LeDoux, J. (1996). *The Emotional Brain*. New York, NY: Simon & Schuster.
- Lerdahl, F. (2013). "Musical syntax and its relation to linguistic syntax," in *Language, Music, and the Brain: A Mysterious Relationship*, ed M. A. Arbib (Cambridge, MA: MIT Press), 257–272. doi: 10.7551/mitpress/9780262018104.003.0010
- Lindblom, B. (1983). "Economy of speech gestures," in *The Production of Speech*, ed P. MacNeilage (New York, NY: Springer-Verlag), 217–245. doi: 10.1007/978-1-4613-8202-7_10
- Lindblom, B. (1990). "Explaining phonetic variation: a sketch of the H and H theory," in *Speech Production and Speech Modelling*, eds W. J. Hardcastle and A. Marchal (Dordrecht: Kluwer Academic Publishers), 413–415. doi: 10.1007/978-94-009-2037-8_16
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 35, 1773–1781. doi: 10.1121/1.1918816
- Lindblom, B., and Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *J. Acoust. Soc. Am.* 50, 1166–1179. doi: 10.1121/1.1912750
- Livingstone, S. R., Thompson, W. F., Wanderley, M. M., and Palmer, C. (in press). Common cues to emotion in the dynamic facial expressions of speech and song. *Q. J. Exp. Psychol.* doi: 10.1080/17470218.2014.971034
- Loucks, T. M., Ofori, E., Grindrod, C. M., De Nil, L. F., and Sosnoff, J. J. (2010). Auditory motor integration in oral and manual effectors. *J. Mot. Behav.* 42, 233–239. doi: 10.1080/00222895.2010.492723
- Madison, G. (2000). Properties of expressive variability patterns in music performances. *J. New Music Res.* 29, 335–356. doi: 10.1080/09298210008565466
- Malécot, A. (1955). An experimental study of force of articulation. *Stud. Linguistica.* 9, 35–44. doi: 10.1111/j.1467-9582.1955.tb00515.x
- McPherson, A. (2013). "Portable measurement and mapping of continuous piano gesture," in *Proceedings of the 13th International Conference on New Interfaces for Musical Expression (NIME)* (Seoul), 152–157.
- McPherson, A., and Kim, Y. (2011). "Multidimensional gesture sensing at the piano keyboard," in *Proceedings of the 29th ACM Conference on Human Factors in Computing Systems (CHI)* (Atlanta, GA), 2789–2798. doi: 10.1145/1978942.1979355
- McPherson, A., and Kim, Y. (2013). "Piano technique as a case study in expressive gestural interaction," in *Music and Human-Computer Interaction*, eds S. Holland, K. Wilkie, P. Mulholland, and A. Seago (London: Springer), 123–138. doi: 10.1007/978-1-4471-2990-5_7
- Minetti, A. E., Ardigo, L. P., and McKee, T. (2007). Keystroke dynamics and timing: accuracy, precision and difference between hands in pianist's performance. *J. Biomech.* 40, 3738–3743. doi: 10.1016/j.jbiomech.2007.06.015
- Morton, E. W. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Amer. Nat.* 111, 855–869. doi: 10.1086/283219
- Munhall, K. G., Ostry, D. J., and Parush, A. (1985). Characteristics of velocity profiles of speech movements. *J. Exp. Psychol.* 11, 457–474. doi: 10.1037/0096-1523.11.4.457
- Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biol. Cybern.* 46, 135–147. doi: 10.1007/BF00339982
- Neuhaus, H. (1973). *The Art of Piano Playing*. New York, NY: Praeger Publishers.
- Ostry, D. J., and Munhall, K. G. (1985). Control of rate and duration of speech movements. *J. Acoust. Soc. Am.* 77, 640–648. doi: 10.1121/1.391882
- Ostry, D., Keller, E., and Parush, A. (1983). Similarities in the control of speech articulators and the limbs: kinematics of tongue dorsum movement in speech. *J. Exp. Psychol.* 9, 622–636. doi: 10.1037/0096-1523.9.4.622
- Padgett, J. (2009). Systemic contrast and Catalan rhotics. *Linguist Rev.* 26, 431–463. doi: 10.1515/tlir.2009.016
- Paeschke, A., Kienast, M., and Sendlmeier, W. F. (1999). "F0-contours in emotional speech," in *Proceedings International Congress of Phonetic Sciences* (San Francisco, CA), 929–932.
- Palmer, C. (2006). "The nature of memory for music performance skills," in *Music, Motor Control and the Brain*, eds E. Altenmüller, M. Wiesendanger, and J. Kesselring (Oxford: Oxford University), 39–53. doi: 10.1093/acprof:oso/9780199298723.003.0003
- Palmer, C., Carter, C., Koopmans, E., and Loehr, J. D. (2007). "Movement, planning, and music: motion coordinates of skilled performance," in *Proceedings of the International Conference on Music Communication Science*, eds E. Schubert, K. Buckley, R. Elliott, B. Koboroff, J. Chen, and C. Stevens (Sydney, NSW: University of New South Wales), 119–122.
- Palmer, C., Koopmans, E., Loehr, J., and Carter, C. (2009). Movement-related feedback and temporal accuracy in clarinet performance. *Music Percept.* 26, 439–450. doi: 10.1525/mp.2009.26.5.439
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York, NY: Oxford University Press.
- Panksepp, J., and Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emot. Rev.* 3, 387–396. doi: 10.1177/1754073911410741
- Parncutt, R., Sloboda, J. A., Clarke, E. F., Raekallio, M., and Desain, P. (1997). An ergonomic model of keyboard fingering for melodic fragments. *Music Percept.* 14, 341–382. doi: 10.2307/40285730
- Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford: Oxford University Press.
- Perkell, J. S., Zandipour, M., Matthies, M. L., and Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *J. Acoust. Soc. Am.* 112, 1627–1641. doi: 10.1121/1.1506369
- Rainville, P., Bechara, A., Naqvi, N., and Damasio, A. R. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *Int. J. Psychophysiol.* 61, 5–18. doi: 10.1016/j.ijpsycho.2005.10.024
- Repp, B. H. (1992a). Diversity and commonality in music performance: an analysis of timing microstructure in Schumann's "Träumerei." *J. Acoust. Soc. Am.* 92, 2546–2568. doi: 10.1121/1.404425
- Repp, B. H. (1992b). A constraint on the expressive timing of a melodic gesture: evidence from performance and aesthetic judgment. *Music Percept.* 10, 221–242. doi: 10.2307/40285608
- Repp, B. H. (1994a). Relational invariance of expressive microstructure across global tempo changes in music performance: an exploratory study. *Psychol. Res.* 56, 269–284. doi: 10.1007/BF00419657
- Repp, B. H. (1994b). On determining the basic tempo of an expressive music performance. *Psychol. Music* 22, 157–167. doi: 10.1177/0305735694222005
- Repp, B. H. (1995). Expressive timing in Schumann's "Träumerei": an analysis of performances by graduate student pianists. *J. Acoust. Soc. Am.* 98, 2413–2427. doi: 10.1121/1.413276
- Repp, B. H. (1996). The dynamics of expressive piano performance: Schumann's "Träumerei" revisited. *J. Acoust. Soc. Am.* 100, 641–650. doi: 10.1121/1.415889
- Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714
- Sandor, G. (1981). *On Piano Playing: Motion, Sound and Expression*. New York, NY: Schirmer Books.
- Scherer, K. R. (1979). "Nonlinguistic vocal indicators of emotion and psychopathology," in *Emotions in Personality and Psychopathology*, ed C. E. Izard (New York, NY: Plenum), 493–529.
- Scherer, K. R. (1995). Expression of emotion in voice and music. *J. Voice* 9, 235–248. doi: 10.1016/S0892-1997(05)80231-0
- Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Commun.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5
- Schwartz, G. E., Weinberger, D. A., and Singer, J. A. (1981). Cardiovascular differentiation of happiness, sadness, anger, and fear following imagery and

- exercise. *Psychosom. Med.* 43, 343–364. doi: 10.1097/00006842-198108000-00007
- Seifert, U., Verschure, P., Arbib, M., Cohen, A., Fogassi, L., Fritz, T., et al. (2013). “Semantics of internal and external worlds,” in *Language, Music, and the Brain: A Mysterious Relationship*, ed M. A. Arbib (Cambridge, MA: MIT Press), 203–232.
- Stein, R. B. (1982). What muscle variables does the central nervous system control? *Behav. Brain Sci.* 5, 535–578. doi: 10.1017/S0140525X00013327
- Steinbeis, N., and Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension resolution patterns. *Cereb. Cortex* 18, 1169–1178. doi: 10.1093/cercor/bhm149
- Sundberg, J. (1982). “Speech, song, and emotions,” in *Music, Mind, and Brain: The Neuropsychology of Music*, ed M. Clynes (New York, NY: Plenum Press), 137–149.
- Sundberg, J. (2000). Four years of research on music and motion. *J. New Music Res.* 29, 183–185. doi: 10.1076/jnmr.29.3.183.3089
- Tortora, G. J. (2002). *Principles of Human Anatomy*. New York, NY: Wiley.
- Uchanski, R. M. (2008). “Clear speech,” in *The Handbook of Speech Perception*, eds D. B. Pisoni and R. E. Remez (Oxford, UK: Blackwell), 207–235.
- van Vugt, F. T., Furuya, S., Vauth, H., Jabusch, H.-C., and Altenmüller, E. (2014). Playing beautifully when you have to be fast: spatial and temporal symmetries of movement patterns in skilled piano performance at different tempi. *Exp. Brain Res.* 232, 3555–3567. doi: 10.1007/s00221-014-4036-4
- Ververidis, D., and Kotropoulos, C. (2006). Emotional speech recognition: resources, features, and methods. *Speech Commun.* 48, 1162–1181. doi: 10.1016/j.specom.2006.04.003
- Westbury, J. R., and Keating, P. A. (1986). On the naturalness of stop consonant voicing. *J. Linguist.* 22, 145–166. doi: 10.1017/S0022226700010598
- Widmer, G., and Goebel, W. (2004). Computational models of expressive music performance: the state of the art. *J. New Music Res.* 33, 203–216. doi: 10.1080/0929821042000317804
- Williams, P. L., and Warwick, R. (1980). *Gray’s Anatomy*. Edinburgh: Churchill Livingstone.
- Wilson, G. D. (1994). *Psychology for Performing Artists: Butterflies and Bouquets*. London: Jessica Kingsley.
- Winges, S. A., Furuya, S., Faber, N. J., and Flanders, M. (2013). Patterns of muscle activity for digital coarticulation. *J. Neurophysiol.* 110, 230–242. doi: 10.1152/jn.00973.2012
- Xu, Y. (2014). *ProsodyPro.praat*. London: University College London.
- Xu, Y., Kelly, A., and Smillie, C. (2013a). “Emotional expressions as communicative signals,” in *Prosody and Iconicity*, eds S. Hancil and D. Hirst (Amsterdam: John Benjamins Publishing Company), 33–60. doi: 10.1075/ill.13.02xu
- Xu, Y., Lee, A., Wu, W.-L., Liu, X., and Birkholz, P. (2013b). Human vocal attractiveness as signalled by body size projection. *PLoS ONE*, 8:e62397. doi: 10.1371/journal.pone.0062397
- Xu, Y., and Wang, M. (2009). Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin. *J. Phon.* 37, 502–520. doi: 10.1016/j.wocn.2009.08.003
- Zachar, P., and Ellis, R. D. (2012). *Categorical Versus Dimensional Models of Affect: A Seminar on the Theories of Panksepp and Russell. Consciousness & Emotion*. Amsterdam: John Benjamins Publishing Company.
- Zanon, P., and De Poli, G. (2003a). Estimation of parameters in rule systems for expressive rendering in musical performance. *Comput. Music J.* 27, 29–46. doi: 10.1162/01489260360613326
- Zanon, P., and De Poli, G. (2003b). Time-varying estimation of parameters in rule systems for music performance. *J. New Music Res.* 32, 295–315. doi: 10.1076/jnmr.32.3.295.16869

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Liu and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.