

Mandarin Chinese pronunciation skills and musical aptitude

JASON KAO

In tone languages such as Mandarin Chinese and Vietnamese, word meaning is not only determined by lexical pronunciation but it is also determined by the manipulation of pitch in every character. In Mandarin Chinese, there are four ways in which pitch can be manipulated, called tones: The first tone is high and level; the second starts low and moderately rises; the third falls and then rises again; the fourth starts high, but drops sharply to the bottom of the tonal range. This study aims at verifying if the production of lexical Mandarin tones varies as a function of musical aptitude and musical exposure. Thirty-one students in their third-year of taking Mandarin at Stuyvesant High School completed a survey for musical exposure and a musical battery for musical aptitude. They then submitted recordings of their voice to undergo pitch analysis process to measure their ability to pronounce Mandarin tones. Initial results show a moderate positive correlation between one's aptitude in tuning (as measured by PROMS subtest 5) and one's ability to pronounce Mandarin tones.

I. INTRODUCTION

Pitch plays an important role in both speech and music. Every language makes use of pitch to convey information about speaker stance, tonality, and emotion. However, in a class of languages known as tone languages, words assume different meanings depending on the lexical tones in which they are spoken. The syllable /ma/ can be associated with at least four different unrelated meanings based on the tone used (Creel, Weng, Fu, Heyman, & Lee, 2017). Mandarin Chinese has four tones manifested by distinctive pitch contours (Figure 1.1). The first tone is high and level; the second starts low and moderately rises; the third falls and then rises again; the fourth starts high, but drops sharply to the bottom of the tonal range (Chun, Jiang, & Avila, 2012). Pitch is also a vital property of music, as it conveys information about tonality (the character of the piece as determined by which key it is played in) (Krumhansl, 1990). These pitch contours (also known as F₀ contours) are one of the most important properties of speech for conveying information about the speaker's intended tonality (character) and emotional state. (T. Kumar, R. Kumar, & V. Kumar, 2013)

Given the importance of pitch and tone for human communication, it is intriguing to note the degree to which the pitch and overall musical ability of individuals vary in their ability to produce tones in tone languages such as Mandarin Chinese.

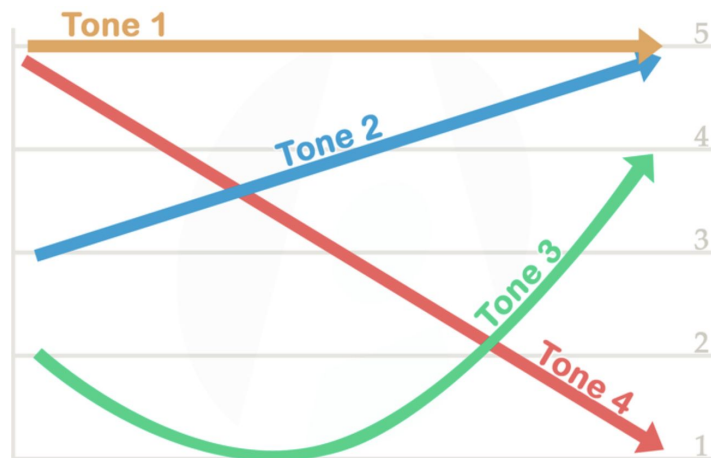


Figure 1.1a. A textbook depiction of the four tones of Mandarin

II. LITERATURE

The ability to discriminate between different pitches of music seems to be related to second language (L2) pronunciation. Milovanov et al. (2010) compared the relationship between foreign language pronunciation and musical aptitude with Finnish learners of English. The subjects were tested on production of English phonemes and a discrimination task of phonemic minimal pairs. (Phonemic awareness is a phonological awareness in which listeners are able to identify between phonemes, the smallest units of sound that can differentiate meaning (Sensenbaugh, 1996)). Milovanov used phonemic minimal pairs for pitch discrimination and chord discrimination tests. However, he found no connection between the performance on English phoneme discrimination tests and English phonemic production ability. However, performance on the English pronunciation test was better for subjects with musical aptitude rather than those with less.

Pfordresher & Brown (2009) also explored whether or not an individual's ability to discriminate and imitate (through singing) musical pitch was dependent on the use of pitch in the individual's native language. However, this study differed in that it specifically compared musical pitch ability of native English speakers to that of native speakers of a tone language (Mandarin, Chinese, Vietnamese). Musicality was measured by pitch perception and pitch production (singing) performance. Pfordresher & Brown discovered that tone language speakers, in comparison to non-tone language speakers, were significantly more accurate at imitating musical pitch and discriminating intervals of pitch.

People with greater abilities to create and detect melody differences also seem to perform better in L2 pronunciation. Delogu, Lampis, & Olivetti (2006) aimed to verify if the discrimination of lexical Mandarin tones varied with melodic ability. They presented subjects with no prior experience of any tone language with two lists of monosyllabic Mandarin words. The students were to identify whether or not the variation between two paired monosyllabic Mandarin words was phonological (in meaning) or tonal (in tone). Subjects who had a higher melodic ability were found to have performed significantly better in variation identification than their less musically talented counterparts.

Part of the reason musical training is beneficial is because of neurological reasons. Milovanov et al. (2008) examined the neurological relationship between musical aptitude and second language pronunciation skills. They found that children with advanced English

pronunciation skills had overall better musical skills than those with less accurate English pronunciation skills. (Overall musicality were measured with the Seashore musicality battery; developed in 1939, it is one of the first musical skill batteries with subtests for musical categories to be created (Lienhard, 1988)). The individual Seashore subtests indicated that the participants with advanced pronunciation skills were superior to the participants with less-advanced pronunciation skills in pitch discrimination ability and sense of tonality.

Moreno et al. (2000) conducted a study to investigate brain plasticity and the effects of short periods of training. Over a six month period, thirty-two non musician children were trained in several aspects of music (e.g. rhythm, melody, harmony). At the end of the training period, the children had significantly improved linguistic pitch processing. Moreno and his team's training these children for a relatively short period had strong consequences on the linguistic organization of the children's brain, demonstrating the fact that pitch processing in language is related to musical training.

III. THEORY

Based on Posedel et al. (2011), in which a greater number of years of musical experience was found to be significantly correlated with a higher pitch perception and pitch perception was found to be a significant predictor of L2 pronunciation quality, there should be a positive relationship between the number of years a subject has played an instrument and his or her Mandarin tone pronunciation performance.

L2 production ability should also increase with current musical training (measured with hours practiced per week), as even six-month-long periods of musical training may significantly improve linguistic pitch processing (Moreno et al., 2000). In tandem with total hours of musical training per week, the type of instrument played should also be a factor of L2 pitch processing. Instruments which require fine tuning (violin, guitar, harp) require their players to pay more attention to closely the tone and pitch of their instrument. Most likely, subjects who play instruments that require fine tuning will perform better in L2 pronunciation than subjects who play instruments that do not require fine tuning. Furthermore, a subject's having absolute pitch (the ability to identify or recreate a musical note without the benefit of an external reference), according to Deutsch, should significantly increase their ability to pronounce tonal languages. (Deutsch, 2013)

Regular exposure to music should also facilitate pitch processing, which is why subjects who listen to music on a regular basis and are involved in a professional group (band, orchestra, DJ-ing) should have perform better in L2 pronunciation than those who do not. Musical training and experience most likely will have a positive impact on L2 phonological production ability as a result of the effects of musical training on pitch perception ability and as suggested by (Posedel, Emery, Souza, & Fountain, 2011).

Finally, based on a review of the literature in the field, the musical aptitude subtests melody, tuning, and pitch, should all be positive predictors of tone ability. (Milovanov et al., 2010; Delogu, Lampis, & Olivetti, 2006)

IV. DATA COLLECTION

This study sought determine whether or not L2 learners with a greater musical background and aptitude would improve L2 production of tones.

Participants

Thirty-one students who were taking their third year of Mandarin at Stuyvesant High School volunteered to participate. A list of the questions in the demographic survey can be found in Appendix A. All of the students had been learning Mandarin from one teacher, Ms. Hui Zhu.

Materials and Procedure

There were two parts to the data collection process. The first was the measuring of musical aptitude. To measure the participants' levels of musical aptitude, participants were given the Profile of Musical Perception Skills (PROMS), a musical test battery which measures perceptual musical skills objectively across multiple domains (melody, tuning, and pitch were tested). Each of the domains' tests contained ten questions. The questions all consisted of three stimuli, with the first two being the same (standard stimulus), and the third being either the same or different (comparison stimulus). Participants were asked to choose on a Likert scale ("definitely different", "I don't know", "probably different", "probably same") how the standard and comparison stimuli compared. (Law & Zentner, 2012)

Based on prior research, the subtests of the PROMS that would most likely predict Mandarin pronunciation ability were melody, tuning, and pitch.

Melody. All melody stimuli were monophonic and composed of a constant rhythm of eight notes. The difficulty of the trials was determined by note density and level of atonality (atonal music is music that seems to lack a clearly defined center ("Atonal")). Figure 3.1 shows examples of easy and complex melody structures. *Tuning.* Each tuning stimulus consisted of C4, E4, G4, and C5, forming a chord of 1.5 seconds in length. The difficulty levels of the stimuli were varied by subtle manipulations to the E4 note. *Pitch.* The pitch subtest played several notes at a constant frequency and intensity. The difficulty level of this subtest was manipulated by varying the degree of pitch difference between the standard and comparison stimulus. (Law & Zentner, 2012)

Example of Easy-Different Trial



Example of Complex-Different Trial



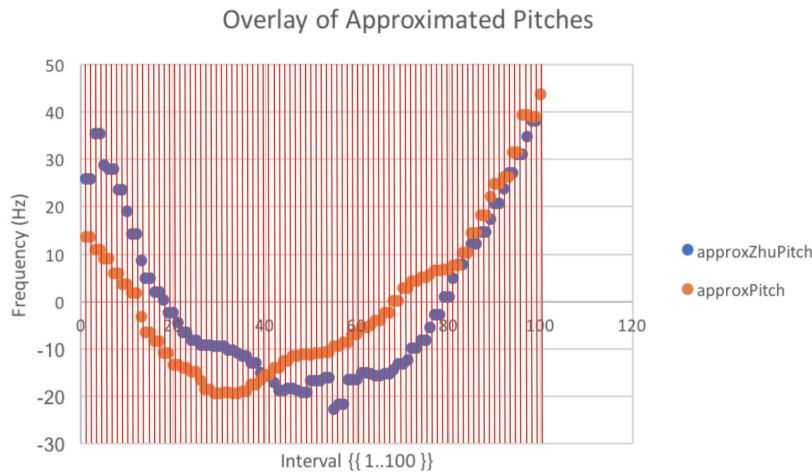
Figure 3.1. Above, an example of an easy comparison trial; below, an example of a complex-different trial. (Law & Zentner, 2012)

The second part of this study was measuring the participants' abilities to pronounce tones. Each was recorded reading a list of four characters, one for each tone: 喝 (hē), 拿(ná), 好 (hǎo), and 不(bù)¹. For each participant's audio file, the pitch contour for each character was

¹ In another study that asked for the recordings of Mandarin students, these four characters were used and provided significant results. (Chun, Jiang, & Avila, 2013)

extracted into an x-variable (time in seconds) and a y-variable (pitch in Hertz) with the acoustic analysis software Praat (version 6.0.28; retrieved on March 23, 2017 from <http://www.praat.org>).

The times and the frequencies of both the subject audio file and the native speaker's audio file were normalized (time was scaled from zero to one, and each frequency value was decreased by the median). Each of the two graphs were approximated into one hundred points (the list of timestamps was transformed into {0.01, 0.02, ..., 0.99, 1.00}). For each approximated point T_i , where T_i is from 0.01 through 1.00, two frequencies (pitches) were taken: the pitch in the subject audio file at time T_i and the pitch in the native speaker's audio file at time T_i . These two pitches were grouped into an (p_i, q_i) coordinate and graphed (Figure 3.2). The R-squared of this plot served as the score for the example user's tone three. This process was conducted for each user's four tones.



The R-squareds of all four analyses of each user were averaged to form a Tone Production Index (TPI).

Figure 3.2. The approximate values for the native speaker's pitches at time T_i and those for the subject's pitches at time T_i were combined into one (p_i, q_i) coordinate and graphed above.

V. RUN THE TEST

To find if musical exposure and musical aptitude were significant predictors of the Tone Production Index (TPI), a multivariate regression was conducted. (The correlation matrix showed no two variables had a strong linear relationship).

	<i>musicalExperi</i>	<i>fineTuning?</i>	<i>practiceTime?</i>	<i>do you listen</i>	<i>sound Engineer</i>	<i>musicality</i>	<i>familyMusicalit</i>	<i>melody</i>	<i>tuning</i>	<i>pitch</i>
<i>musicalExperi</i>	1									
<i>fineTuning?</i>	0.29015959	1								
<i>practiceTime?</i>	0.38336523	0.33524932	1							
<i>How often do</i>	0.04692331	-0.27086817	-0.00990148	1						
<i>Are you invol</i>	0.33836425	-0.15488062	0.33438812	0.01786854	1					
<i>musicality</i>	0.46126928	0.18626902	0.30351385	0.19721973	0.13145556	1				
<i>familyMusical</i>	0.18424789	-0.15510917	0.40978293	-0.0190879	0.38510245	0.44377105	1			
<i>melody</i>	0.55493055	0.21227495	0.10000875	0.05757158	0.02814252	0.28031878	0.12833212	1		
<i>tuning</i>	0.42192287	-0.16761222	-0.05546607	0.18418859	0.10612117	0.03725898	-0.0396374	0.37937474	1	
<i>pitch</i>	0.15141119	-0.28256818	-0.06219321	0.25749272	-0.12582246	0.03538752	0.15574885	-0.09583364	0.48808586	1

Full Regression (the full output can be seen in Appendix C)

<i>p-value</i>	<i>Expected Sign</i>	<i>Coefficient (Standard Error)</i>	<i>Adjusted</i>
intercept		0.34468 (0.13941)	0.011257929
musical experience	positive	-0.00041 (0.00775)	0.479118969
fine tuned instrument?	positive	0.1242 (0.07421)	0.054898947*
practice per week	positive	0.00995 (0.05859)	0.433416219
daily exposure to music	positive	0.00288 (0.02291)	0.450593433
involved in pro group?	positive	-0.12816 (0.13263)	0.172716349
musical status	positive	-0.01547 (0.04738)	0.373700048
family's musical status	positive	-0.00182 (0.07111)	0.489934356
PROMS melody subtest	positive	-0.00884 (0.01004)	0.194713267
PROMS tuning subtest	positive	0.01888 (0.00951)	0.030494889**
PROMS pitch subtest	positive	0.00316 (0.01471)	0.416161095
R-squared	0.2483150550	*, **, *** indicates significant levels at the 90%, 95%, and 99% respectively.	
Adjusted R-squared	0.1946232732		
No. Observations	31		

The model had an R-squared of 0.37998 and an Adjusted R-squared of 0.06998. However, it was found that whether or not a participant played a fine tuned instrument (e.g. violin, guitar) significantly predicted TPI at the 90% significance level ($\beta_1 = 0.12$, $p < 0.10$) as did the participant's score on the PROMS tuning subtest at the 95% significance level ($\beta_1 = 0.02$, $p < 0.05$). The regression output allows us to construct such an equation:

$$\text{TPI} = 0.34468 + -0.00041(\text{Years of Musical Experience}) + 0.1242(\text{Plays Fined Tuned Instrument?}) + 0.00995(\text{Hours of Practice per Week}) + 0.00288(\text{Daily Exposure to Music}) - 0.12816(\text{Involved in Professional Group?}) + -0.01547(\text{Musical Status}) + -0.00182(\text{Family Musical Status}) + -0.00884(\text{PROMS Melody Subtest}) + 0.01888(\text{PROMS Tuning Subtest}) + 0.00316(\text{PROMS Pitch Subtest})$$

However, because of the insignificant F-statistic [$F(10,20) = 1.22572$, $p = 0.33343$], the above regression equation does not have validity in fitting the data.

To better the fit of the model, the variables were downselected. Besides from fine tuned instrument? and the PROMS tuning subtest, the variables which had the smallest p-values in the regression were involvement in professional groups (e.g. band, DJ-ing) and the PROMS melody subtest (as expected). Since both the variables fine tuned instrument? and involvement in a professional group were binary, the line fit plots for both were not helpful in determining their fit-ness. PROMS melody and PROMS tuning line fit plots are plotted in Figure 5.1a and 5.1b, respectively.

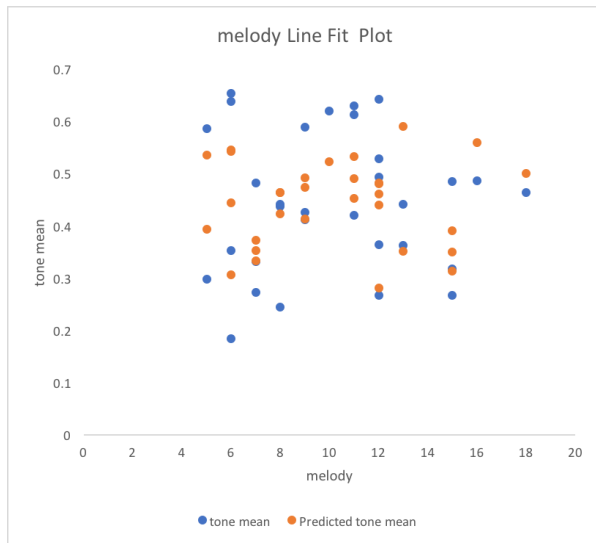


Figure 5.1a. Although the regression output gave PROMS melody score a negative coefficient, it is clear from the line fit plot that there is no real correlation.

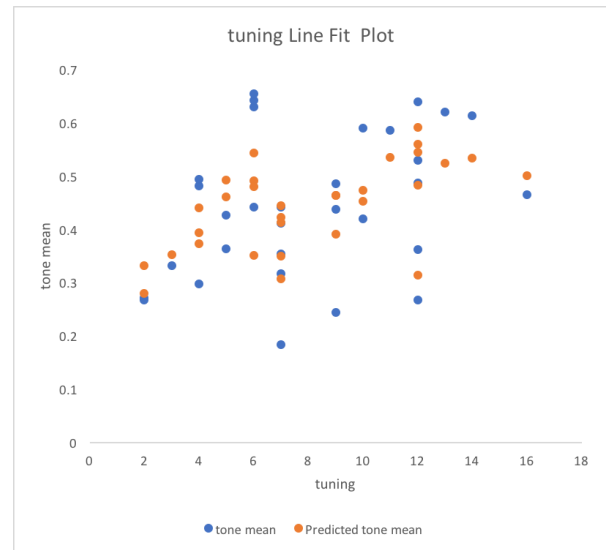


Figure 5.1b. There seems to be a weak positive correlation between PROMS tuning score and there is no real correlation.

Because of PROMS melody's having no correlation with the TPI, it was not chosen in the down selecting. Moreover, although professional group involvement was not shown in Figure 5.1, there were only two out of the thirty-one participants who were enrolled in a profession listening group, rendering the variable useless. In conclusion, the variables which were chosen were whether or not a participant played an instrument which required fine tuning and the score a participant received on the tuning subtest of the PROMS.

Downselected Regression (the full output can be found in Appendix D)

	Expected Sign	Coefficient (Standard Error)	Adjusted
<i>p-value</i>			
intercept		0.29434 (0.05678)	1.67592E-05
fineTuning?	positive	0.10686 (0.05032)	0.021352362**
tuning	positive	0.01517 (0.00606)	0.009221139***
R-squared	0.37998361	*, **, *** indicates significant levels at the 90%, 95%, and 99% respectively.	
Adjusted R-squared	0.069975415		
No. Observations	31		

After down selecting the variables of the model, the fit of the model increased significantly [$F(2,28) = 4.62482$, $p = 0.01839$]. There was no pattern in the residuals (Figure 5.2). Whether or not a participant played a fine tuned instrument became a more significant predictor ($\beta_1 = 0.11$, $p < 0.05$) as did tuning ($\beta_1 = 0.015$, $p < 0.01$). Therefore, at a 95% significance level, there is strong statistical evidence that this model has validity in fitting the data. A new equation for TPI can now be written:

$$\text{TPI} = 0.29434 + 0.10686(\text{plays Fine Tuned Instrument?}) + 0.015172(\text{PROMSTuningSubset})$$

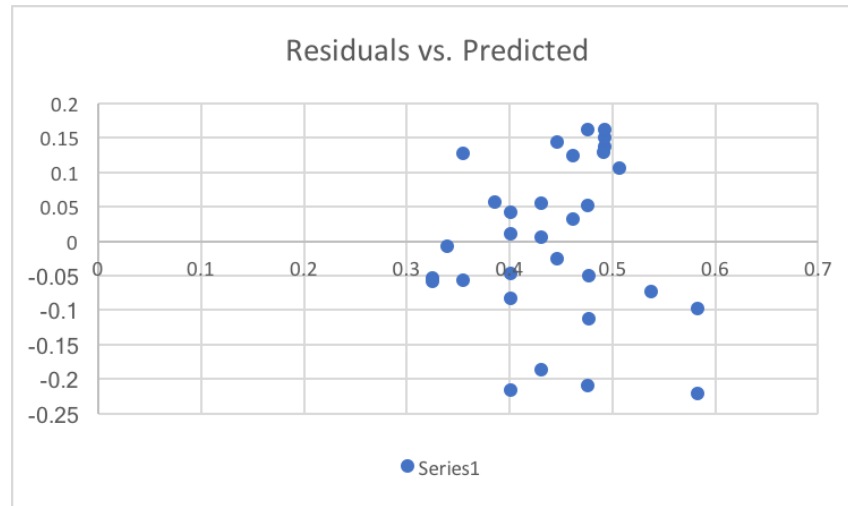


Figure 5.2. The residuals of the downselected regression contain no pattern within these residuals.

VI. DISCUSSION

This paper presented an analysis of pitch contours and musical aptitude with the goal of finding a positive correlation between musical aptitude and the ability of a Mandarin student to pronounce tones in Mandarin. In the study, a multivariate regression was conducted between several variables. It was found that whether or not a participant played a fine tuned instrument and musical aptitude in tuning were both significant predictors of Mandarin tone pronunciation skill. Whether or not a participant played a fine tuned instrument was found to be positively correlated with tone pronunciation, and there was strong statistical evidence that the regression model with only those two variables had validity in fitting the data. Finally, the results of the regression suggest that musical experience and amount of practice per week are not significant predictors of Mandarin tone language, despite the research of Milovanov et al. (2008) and Moreno et al. (2000).

There exist several limitations of this current study. The first is in the data collection process. The environment was not controlled for the recording of participant audio files and the completing of PROMS audio tests, in which the differences between stimuli are very sensitive. The fact that some users' environments may have been silent while others may have been very noisy definitely contribute to the quality of the data collected. Moreover, there may exist voluntary bias in the data collection bias. More who are confident in their musical ability took the test than those who are not. (Countless people were curious after taking the test as to what score they received). Furthermore, this study's small sample size of thirty-one was most likely the reason for very few significant variables. A small sample size was also the reason for the randomness as seen in Figure 5.1a.

The findings from this study may be extended into Mandarin classrooms. The fact that tuning is a significant predictor of pronunciation ability benefits those who write curriculums for teaching Mandarin Chinese. If music is incorporated into the Mandarin curriculum (for

example, through class singing, in which everybody must tune with everybody else), then students may be able to pronounce lexical Mandarin tones more naturally.

In conclusion, initial results show trends between the aforementioned variables. But despite a few significant explanatory variables, due to a small sample size and an uncontrolled recording/listening environment, more rigorous analysis is necessary.

VII. Appendix

A. PROMS Demographic Survey Questions

1. Gender (male/female)
2. Age
3. Do you play a musical instrument (yes/no)
4. Do you have perfect pitch? (yes/no)
5. How often do you listen to music? (never, occasionally, 1-2 times/week, 3-4 times/week, 5-6 times/week, everyday)
6. Are you involved actively in professional listening activities? (e.g. Conducting, Sound Engineering, Piano Tuning, Performing, DJ-ing, Music Perception Research and others)
7. Would you consider yourself a(n): non-musician, music-loving non-musician, amateur musician, semi-professional musician, profession musician
8. Are any members of your family musicians? (no, yes--amateurs, yes--professionals)

B. PROMS Survey Responses

C. Full Regression Output

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.61642811							
R Square	0.37998361							
Adjusted R Square	0.06997542							
Standard Error	0.12989235							
Observations	31							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	10	0.20680397	0.0206804	1.22572118	0.33343752			
Residual	20	0.33744048	0.01687202					
Total	30	0.54424445						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	0.34468478	0.13940815	2.47248668	0.02251586	0.05388449	0.63548508	0.05388449	0.63548508
musicalExperience	-0.0004109	0.0077495	-0.0530252	0.95823794	-0.0165761	0.01575426	-0.0165761	0.01575426
fineTuning?	0.12419844	0.07421465	1.67350293	0.10979789	-0.0306106	0.27900749	-0.0306106	0.27900749
practiceTime	0.00995161	0.05858972	0.16985255	0.86683244	-0.1122644	0.13216763	-0.1122644	0.13216763
How often do you practice?	0.00288048	0.02290679	0.12574803	0.90118687	-0.0449022	0.0506632	-0.0449022	0.0506632
Are you involved in a band or orchestra?	-0.1281579	0.13262699	-0.9663034	0.3454327	-0.404813	0.14849714	-0.404813	0.14849714
musicality	-0.0154733	0.04738392	-0.326551	0.7474001	-0.1143144	0.08336786	-0.1143144	0.08336786
familyMusicality	-0.0018169	0.0711094	-0.025551	0.97986871	-0.1501485	0.14651469	-0.1501485	0.14651469
melody	-0.0088353	0.01004258	-0.8797819	0.38942653	-0.0297837	0.01211318	-0.0297837	0.01211318
tuning	0.01887994	0.00950964	1.98534733	0.06098978	-0.0009568	0.0387167	-0.0009568	0.0387167
pitch	0.00315548	0.01471024	0.21450897	0.83232219	-0.0275295	0.03384049	-0.0275295	0.03384049

D. Downselected Regression Output

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.4983122							
R Square	0.2483151							
Adjusted R Square	0.1946233							
Standard Error	0.1208748							
Observations	31							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	2	0.1351441	0.067572	4.6248243	0.0183866			
Residual	28	0.4091004	0.0146107					
Total	30	0.5442444						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	0.2943445	0.0567765	5.1842712	1.676E-05	0.1780432	0.4106458	0.1780432	0.4106458
fineTuning?	0.106856	0.0503264	2.1232608	0.0427047	0.0037671	0.2099449	0.0037671	0.2099449
tuning	0.0151727	0.0060629	2.502558	0.0184423	0.0027535	0.027592	0.0027535	0.027592

E. Oral Presentation Slides