Assessment

1. Automatic Evaluation of Aspects of Performance and Scheduling in Playing the Piano ~ 우리 연구와 매우 흡사함

논문: https://dl.acm.org/doi/10.1145/3503252.3531297

- 피치, 템포, 리듬, 아티큘레이션에 대해서 점수를 자동으로 평가

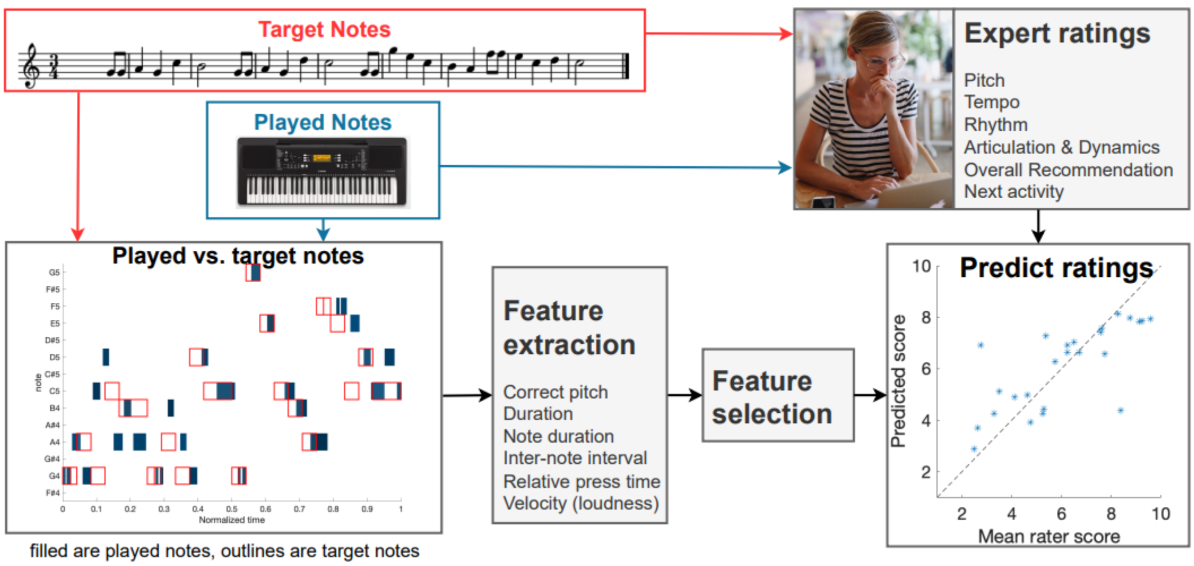
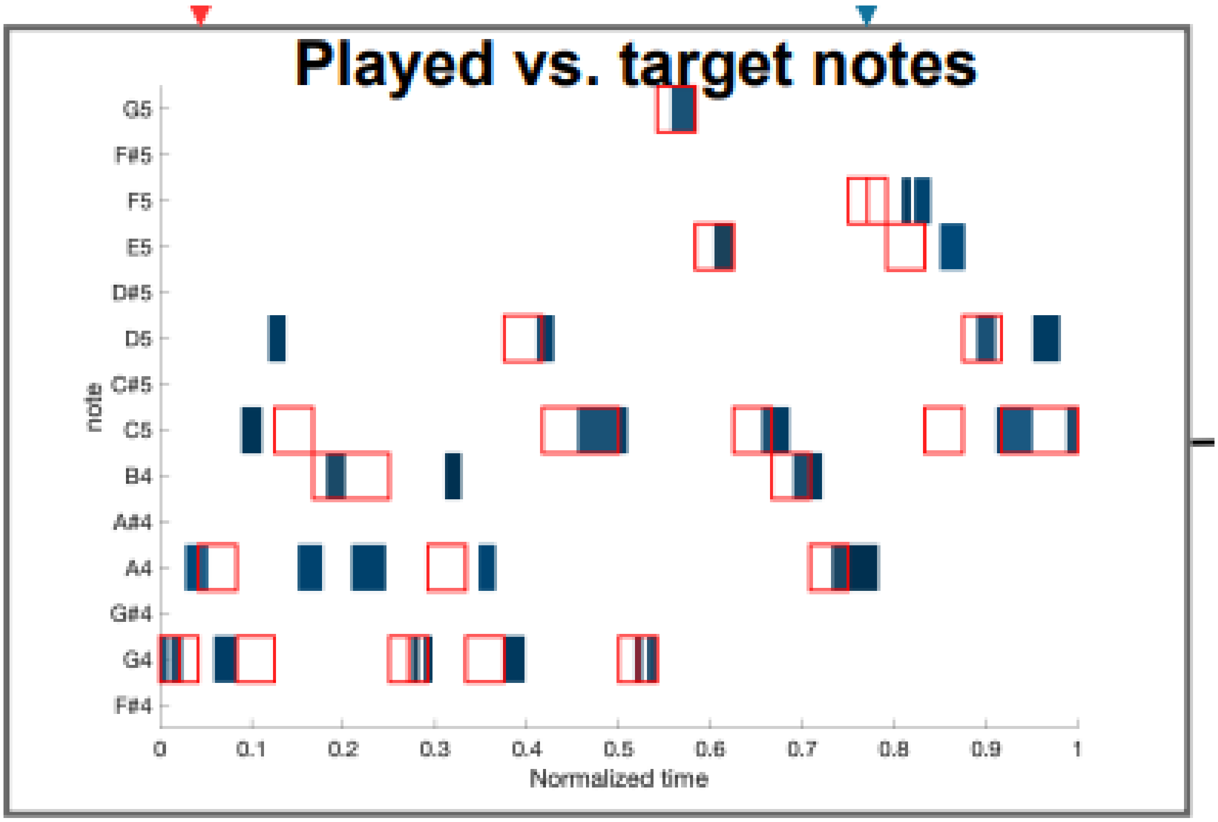


Figure 1: Overview of the process described in the paper: From a set of MIDI recordings, features are extracted based on comparisons to the target notes (musical score). From these features, the most informative are selected, and used to predict ratings of evaluation of different aspects of performance given by experts in the field.

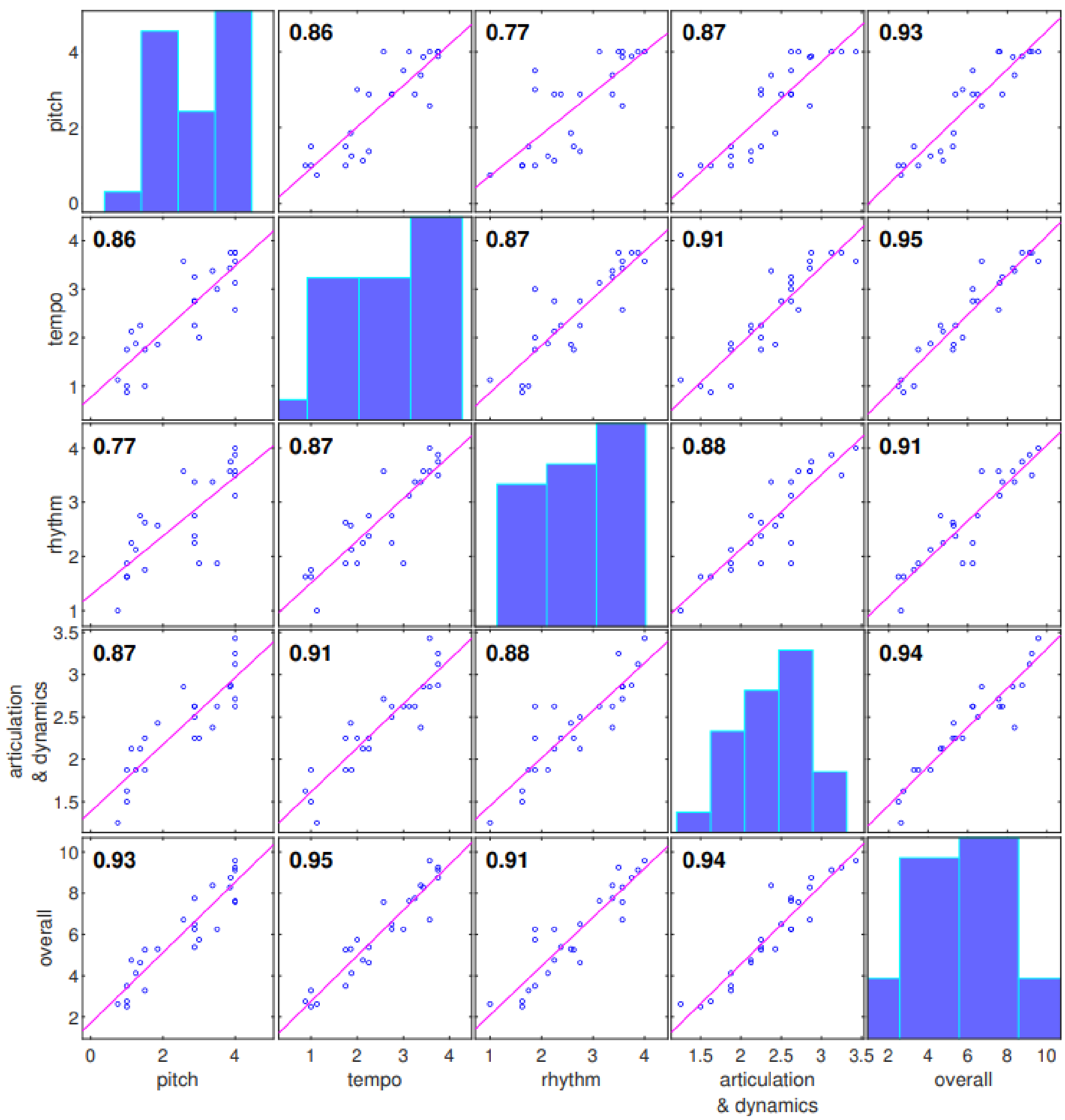


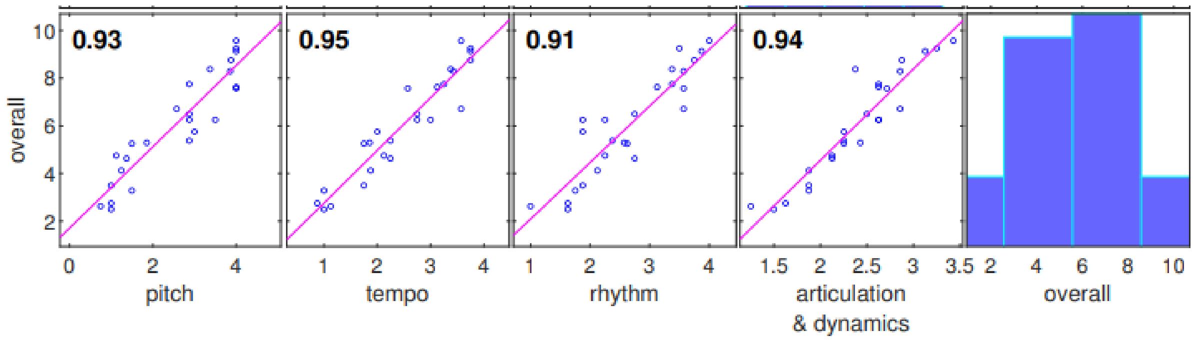
filled are played notes, outlines are target notes

- 특징 추출: Correct pitch, Duration, Note duration, Inter-note interval, Relative press time Velocity (loudness)
- 예측에 사용할 특징을 선택하기 위해 특징별로 라쏘 적합(lasso fit)을 수행하여(4-fold cross validation 사용) 교차 검증 결과 MSE(평균제곱오차)가 가장 작은 값을 선택
- 선택된 특징을 사용하여 선형 회귀를 통해 관련 가중치(weights) 추출 → R^2 measure를 보고 적합도를 결정

Feature name	Pitch	Tempo	Rhythm	Articulation & Dynamics	Overall	Recomm-endation	Same choice	Different choice
Correct pitch	X	X	X	X	X	X	X	X
Duration		X	X	X	X			
Note duration (slope)								
Note duration (mean)						X		
Note duration (std)								
Inter-note interval (slope)								X
Inter-note interval (mean)								
Inter-note interval (std)	X	X		X	X	X	X	
Relative press time (slope)								
Relative press time (mean)		X	X		X			
Relative press time (std)								
Velocity (slope)								
Velocity (mean)								
Velocity (std)								

Table 2: X indicates that a particular features will be used in the linear regression, based on the outcome of the lasso fit, except for recommendation, which was based on chi squared tests





Overall 을 보면 0~10

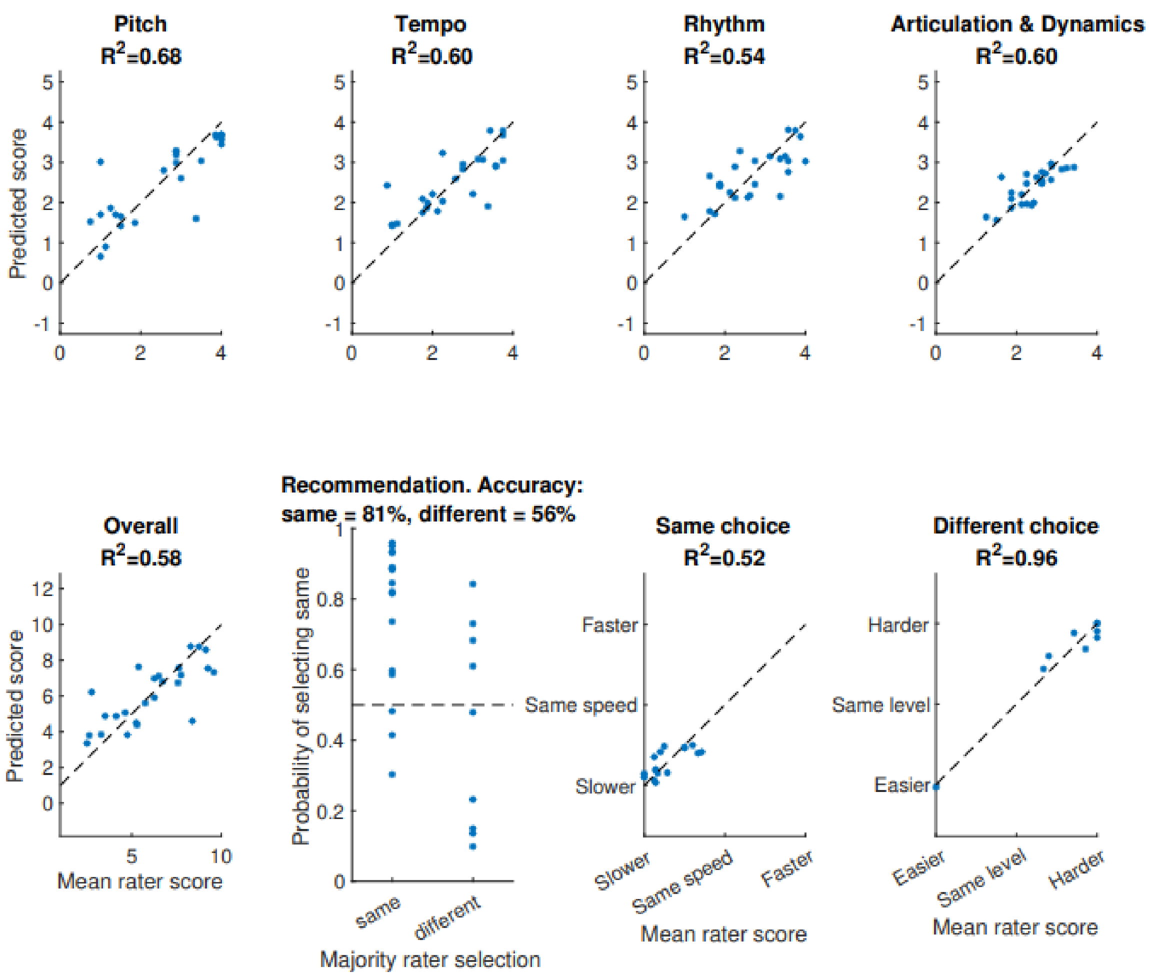


Figure 5: The mean scores of the raters (x-axis) and predicted scores from the linear regression (y-axis)

$$
\begin{aligned}
\text{Pitch} &= +14.55 \times \text{Correct pitch} - 0.70 \times \text{Inter-note interval (std)} - 10.81 \\
\text{Tempo} &= +7.62 \times \text{Correct pitch} - 0.68 \times \text{Duration} \\
&\quad - 0.14 \times \text{Inter-note interval (std)} - 1.96 \times \text{Relative press time (mean)} - 5.54 \\
\text{Rhythm} &= +6.75 \times \text{Correct pitch} - 0.74 \times \text{Duration} \\
&\quad - 1.62 \times \text{Relative press time (mean)} - 4.47 \\
\text{Articulation \& Dynamics} &= +5.43 \times \text{Correct pitch} - 0.37 \times \text{Duration} \\
&\quad - 0.16 \times \text{Inter-note interval (std)} - 2.67 \\
\text{Overall} &= +21.71 \times \text{Correct pitch} - 1.11 \times \text{Duration} \\
&\quad - 0.47 \times \text{Inter-note interval (std)} - 3.19 \times \text{Relative press time (mean)} - 15.70 \\
P(same) &= 1/\left(1 + e^{-(8.81 + -13.11 \text{Correct pitch} + 7.33 \text{Note duration (mean)} + 1.06 \text{Inter-note interval (std)})}\right) \\
\text{Same choice} &= +2.41 \times \text{Correct pitch} - 0.18 \times \text{Inter-note interval (std)} - 0.84 \\
\text{Different choice} &= +10.04 \times \text{Correct pitch} + 25.69 \times \text{Inter-note interval (slope)} - 7.19
\end{aligned}
$$

Table 3: The regression equations calculated from the data

데이터를 통해 얻어진 회귀 방정식

Correct pitch

- 정답 데이터에서 올바르게 연주된 음표의 비율로 정의

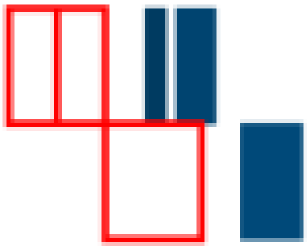
Duration

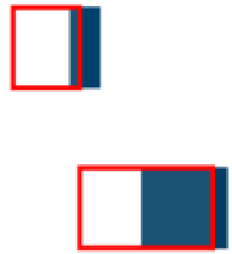
- 특징(duration) = (인풋 데이터 duration - 정답 데이터 duration) / 정답 데이터 duration
    - 악보의 정답 데이터 duration에 맞게 정규화됨

$$\text{Feature}_{\text{duration}} = (\text{Duration}_{\text{actual}} - \text{Duration}_{\text{ideal}})/\text{Duration}_{\text{ideal}}$$

Note duration

- 악보 데이터의 이상적인 note duration - 인풋 데이터의 note duration (즉, 상대적인 차이를 이용) ~ 이를 통해 얼마나 곡의 박자(?)를 잘 유지하고 있는지(Slope, 기울기), 연주 템포가 얼마나 적절한지(Mean, 평균), 얼마나 일관되게 연주하고 있는지(SD, 표준편차)를 추정할 수 있음
    - 다음 그림에서 빨간 테두리 박스가 정답 악보 데이터, 색칠된 박스가 사용자의 입력 데이터





Inter-note intervals(음표 간 간격)

- 현재 노트의 시작 시작(current onset) - 이전 노트의 시작 시간(prev onset)
    - onset: starting time

Relative press time(상대적인 누른 시간)

- 음표 지속 시간 / 음표 간의 간격(current onset - prev onset)
- 이를 활용해서 스타카토, 레가토, 테누토 등을 구할 수 있음
    - 상대적으로 누르는 시간이 짧다면 스타카토, 누르는 시간이 길다면 레가토
    - 예를 들어, "lingering"이라면 이 시간이 1보다 커질 것임

Velocity (loudness) ~ 속도 & 크기

- 평가 기준: 연주된 인풋 곡 미디 데이터와 악보대로 연주한 정답 미디 데이터의 상대적인 속도(velocity)의 차이를 비교

문제점: 악보를 보고 그대로 연주를 지속하는 것은 전문가나 중급자 이상은 가능하지만 초보자의 경우, 연주를 하다가 잘못 쳤을 때 멈추는 경우가 있다는 문제점이 있음

---

2. Use of eigenperformances to analyze proficiency of piano performance

논문: https://www.jstage.jst.go.jp/article/ast/36/4/36_E1408/_pdf

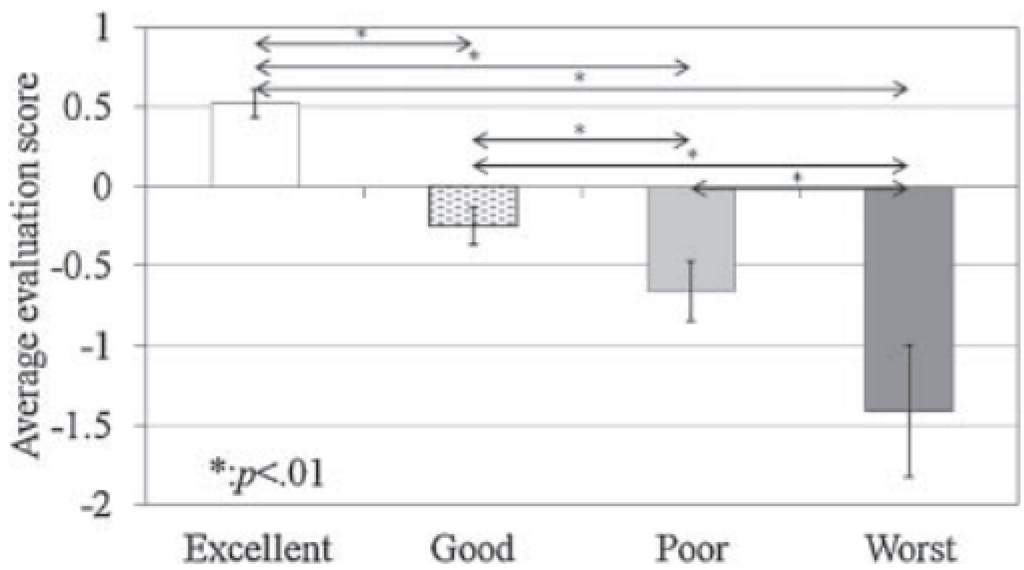


Fig. 1 Average and 95% CI (confidential interval) for cluster of proficiency score.

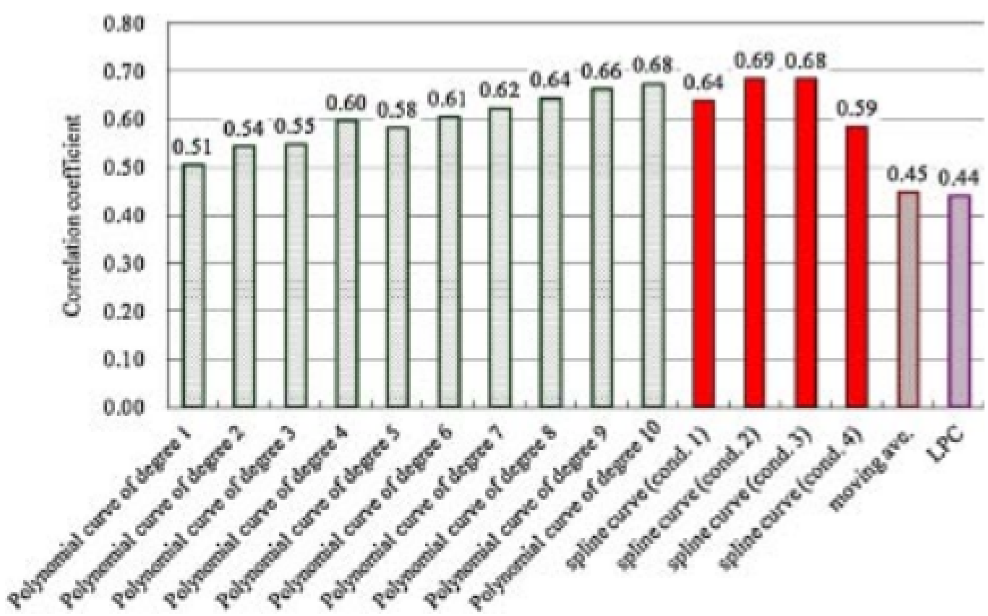


Fig. 2 Correlation coefficients of estimated scores under each condition and evaluation score by piano experts.

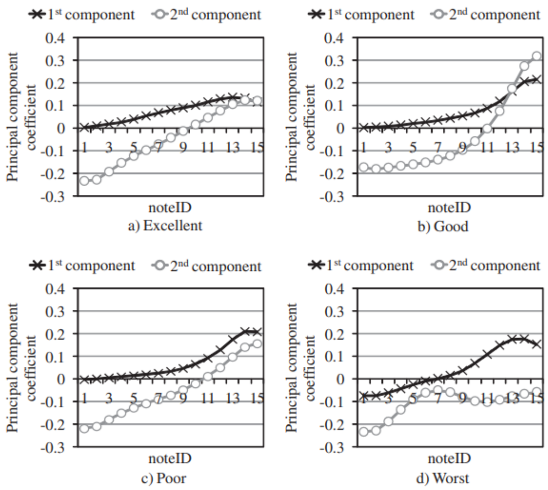


Fig. 3 Eigenperformances for onset.

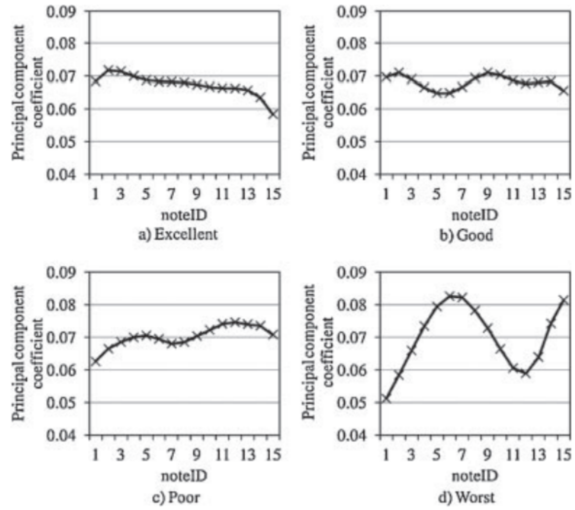


Fig. 4 Eigenperformances for MIDI-velocity.

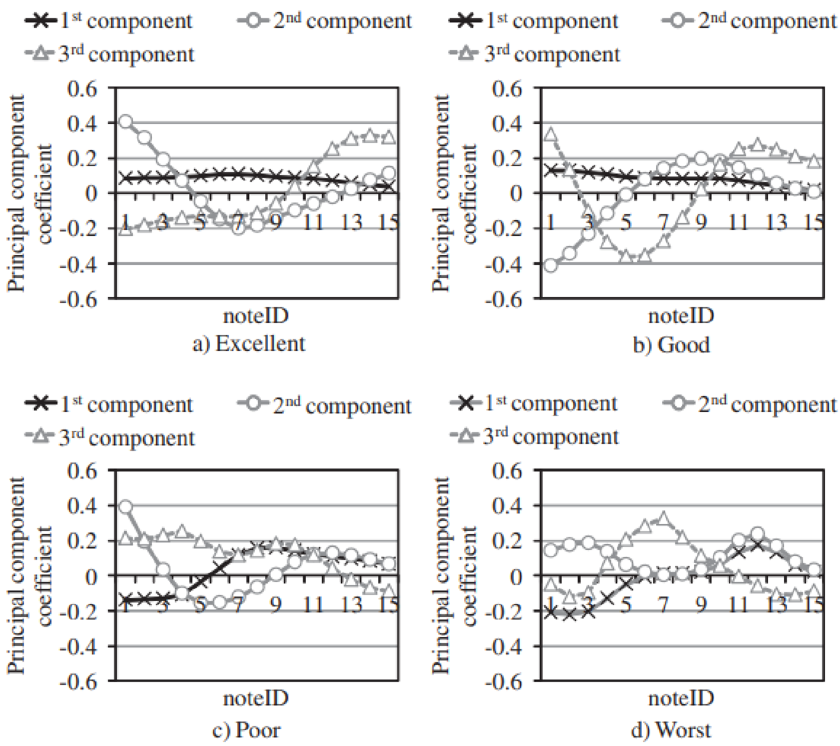


Fig. 5 Eigenperformances for duration.

Table 1 Result by eigenperformances.

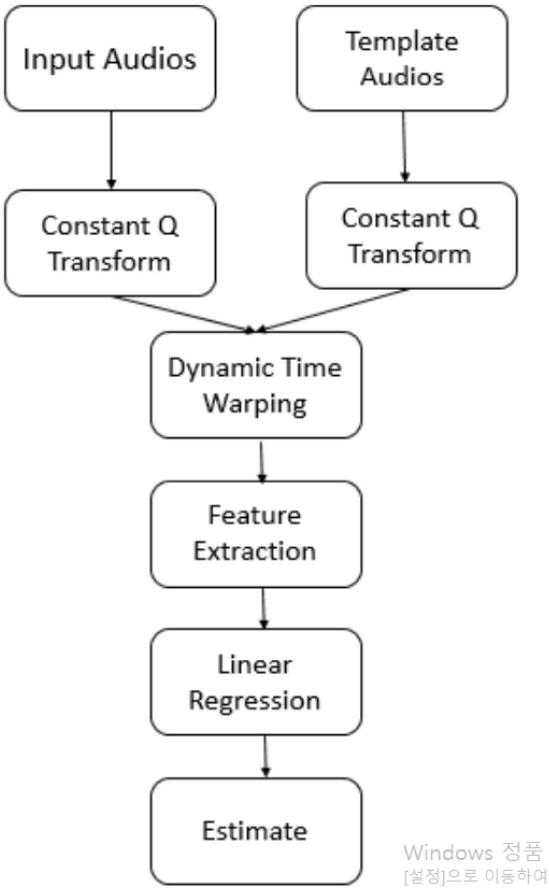
	Onset	MIDI-velocity	Duration
Excellent	Linear change No steep change Constant tempo	Linear change No steep change Constant value	Constant value in 1st PC Nearly straight line in 1st PC
Good	Smooth change	No linear change	Constant value in 1st PC Nearly straight line in 1st PC
Poor	Shape of resembling wavy curve	Slightly wavy change	Steep and unstable change in 2nd and 3rd PC
Worst	Unstable change Steep change	Steep change Wavy change	Steep and unstable change in 2nd and 3rd PC

- K-means Clustering 알고리즘으로 Tempo 평가
    - 마디마다 혹은 프레임 집합(예를 들면, 100개 프레임 집합) 단위로 실제 연주한 velocity와 정답 데이터의 velocity의 변화를 묶어서 clustering 해주면 될 것 같

---

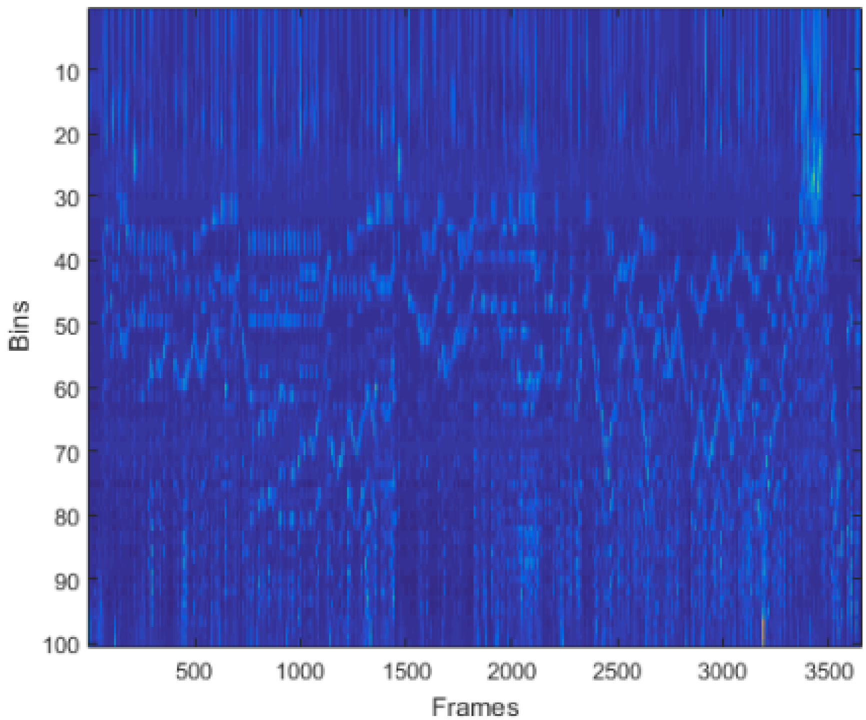
3. An efficient audio based performance evaluation system for computer assisted piano learning

https://ieeexplore.ieee.org/document/8393203

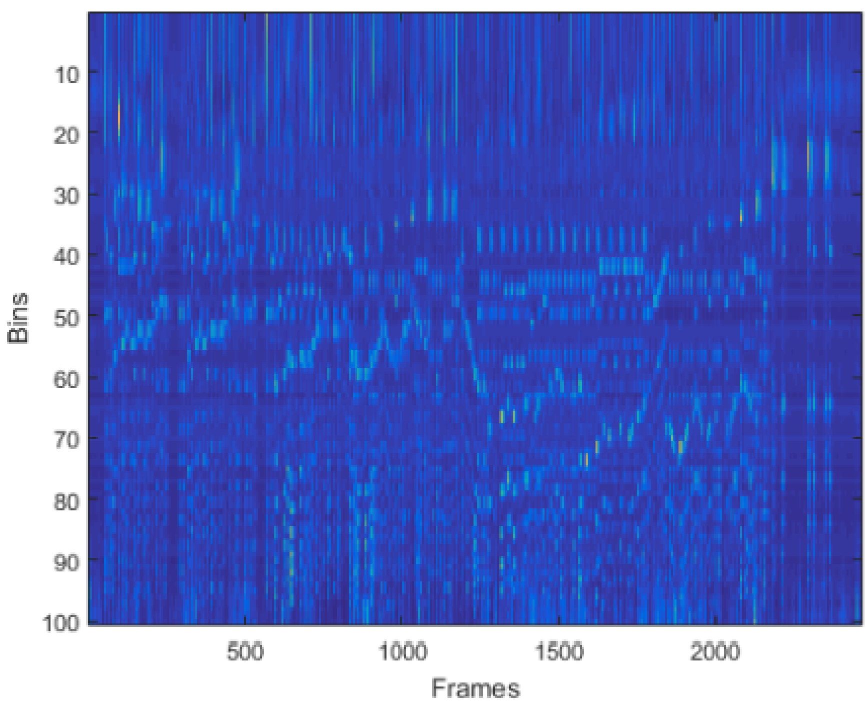


(a) Feature Sequence of Model Audio

- Most of the pianos contains 88 keys. The piano sound could be considered as a set of music notes with certain rhythms.
- CQT is widely applied inmusic processing due to its log-scale property in frequency domain.



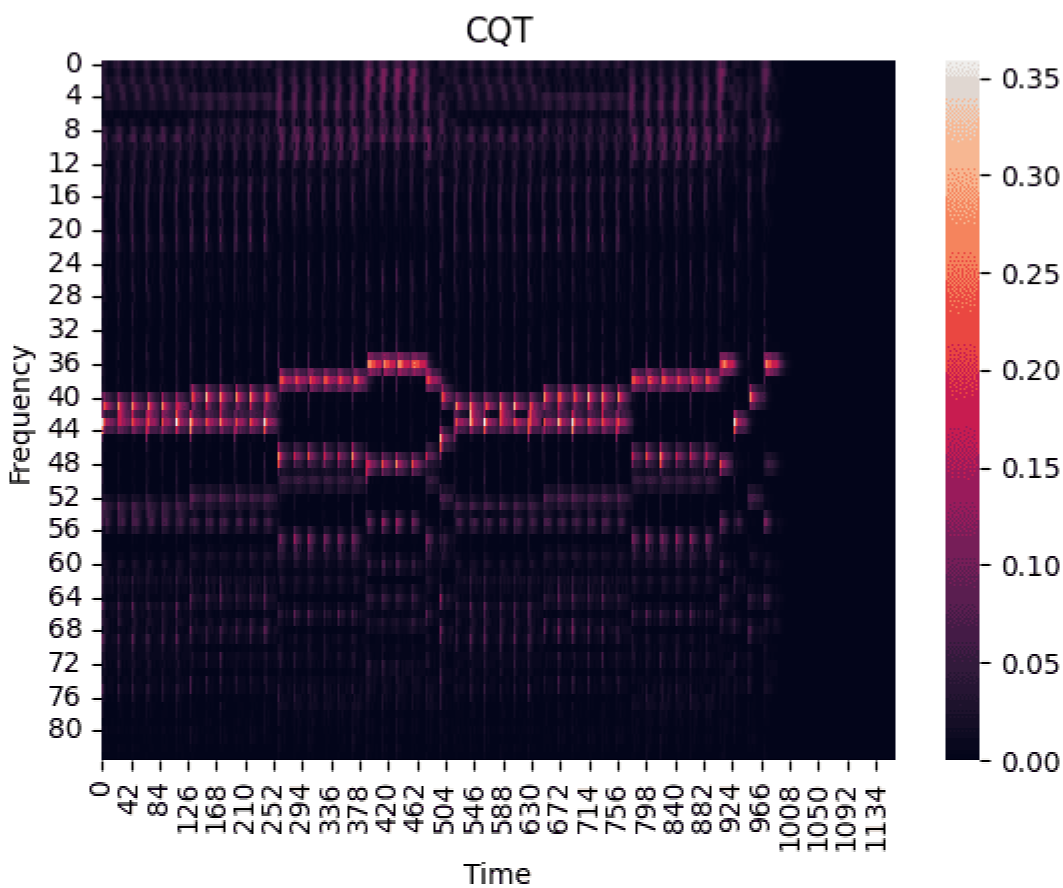
(a) Feature Sequence of One Template Audio



(b) Feature Sequence of Test Audio

Fig. 2. Example of One Feature Sequences

- CQT: https://core.ac.uk/download/pdf/144846462.pdf
    - CQT로 만든 스펙토그램은 음악 오디오 신호 처리에 적합함 / 주파수의 로그 스케일 표현을 사용하는데 이는 12음계의 피치를 커버하는데 적합함
    - CQT에서 시간 해상도는 SR(Sample Rate)와 Hop Length에 의해 결정(일반적으로 사용되는 SR은 44100, 성능을 고려한다면 16000도 괜찮음)
    - 피아노는 총 88개의 건반을 가지므로 264(3 * 88)의 n_bins를 설정하여 주파수 해상도를 설정



- DTW(Dynamic Time Warping): a popular time series analysis method aiming at aligning two temporal sequences with different duration and speed, resulting in finding an optimal alignment. ~ 지속 시간, 속도가 다른 두 시계열 데이터를 비교해서 최적의 정렬을 진행
- 우리 연구의 경우, 정답 데이터와 인풋 데이터의 길이가 다르므로 DTW를 통해 비교하기 전 정렬이 필요

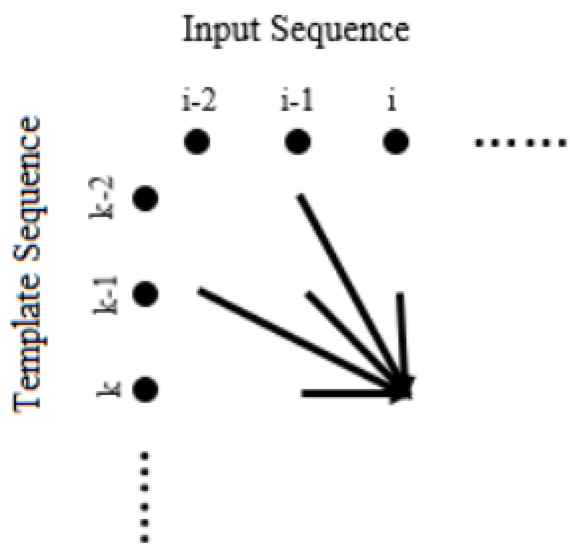


Fig. 3. Scheme of DTW

1. 인풋 데이터와 정답 데이터 간의 연주 시간이 다르기 때문에 DTW로 프레임을 정렬한다.

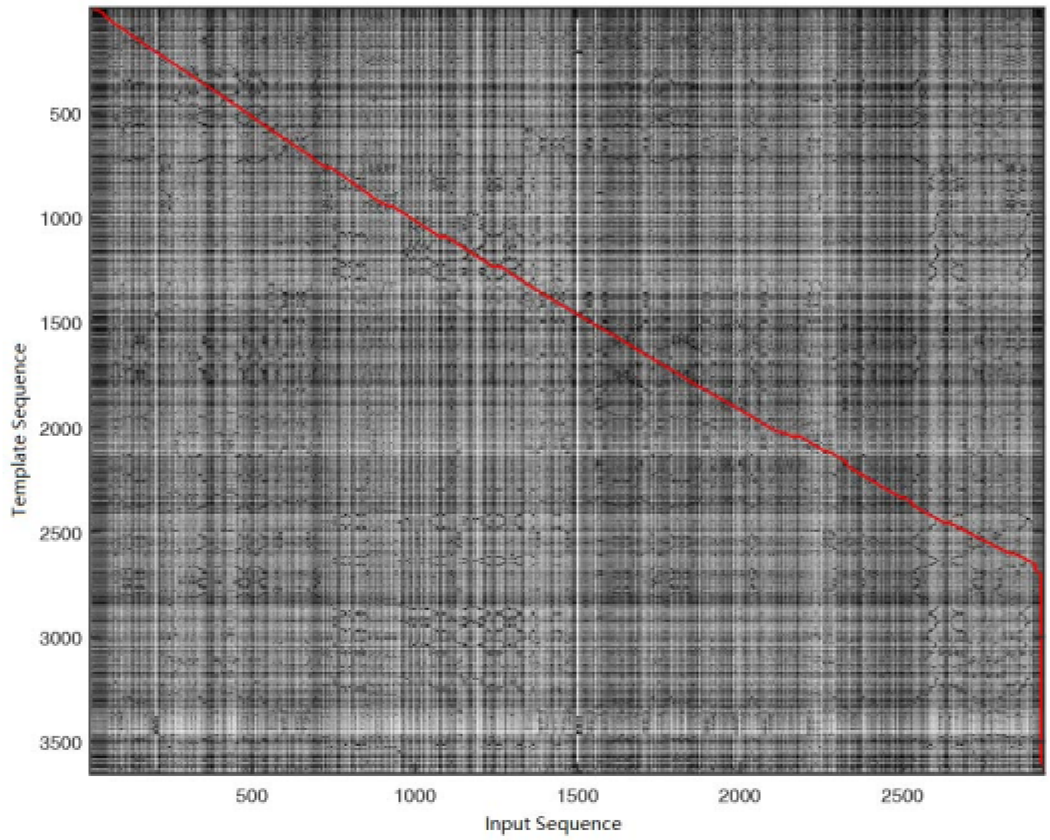
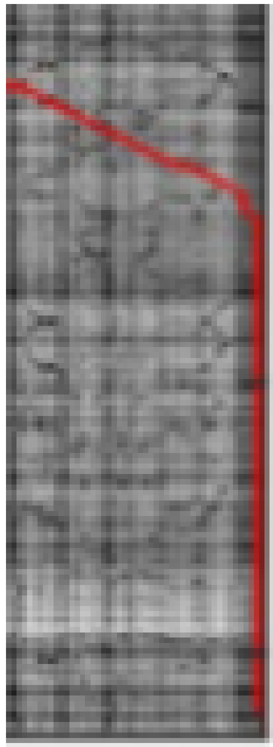


Fig. 4. Example of DTWmatching path

- 위 그림에서 Input Sequence의 2700 프레임 정도를 확인하면 다음과 같이 나타나면서 그 이후 프레임부터는 일치하지 않음을 뜻한다.



- 이후, 3명의 전문가가 연주 평가 점수를 라벨링한 데이터를 최소 제곱법을 활용하여 평가함
- 점수대는 0 ~ 100점으로 평가하고, 60~100점 범위 내에 집중되어 있음

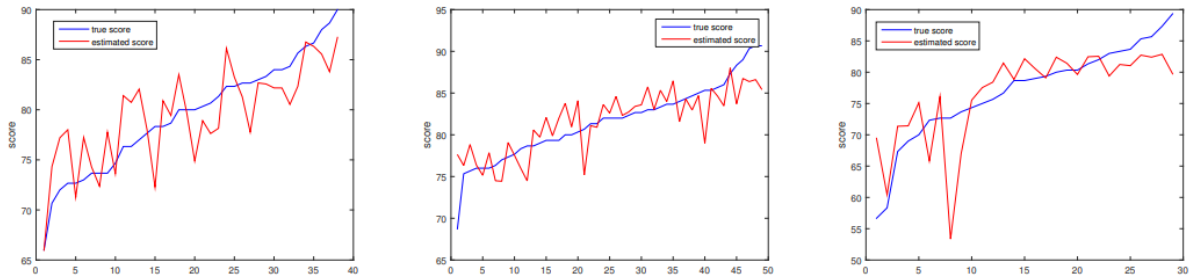


Fig. 5. The estimated score and true score of each song

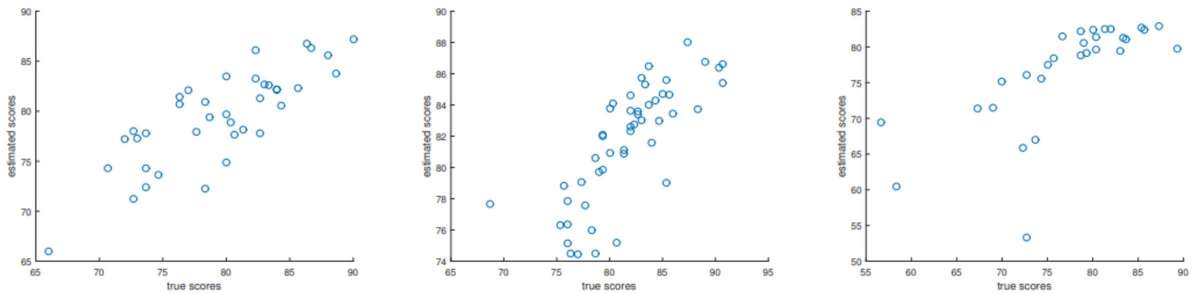
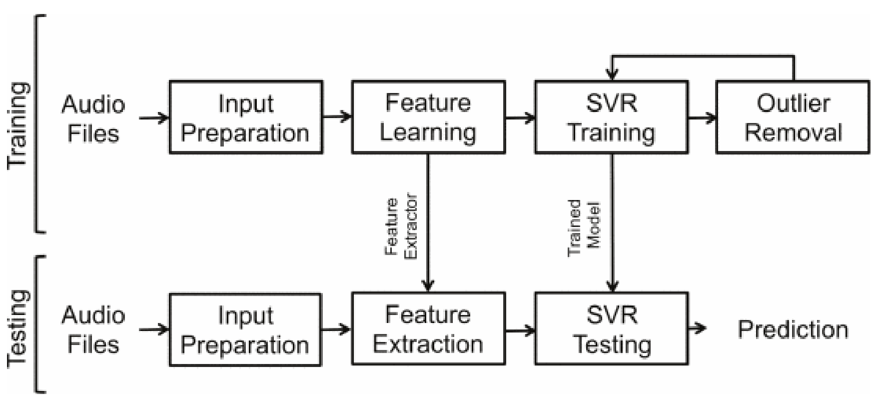
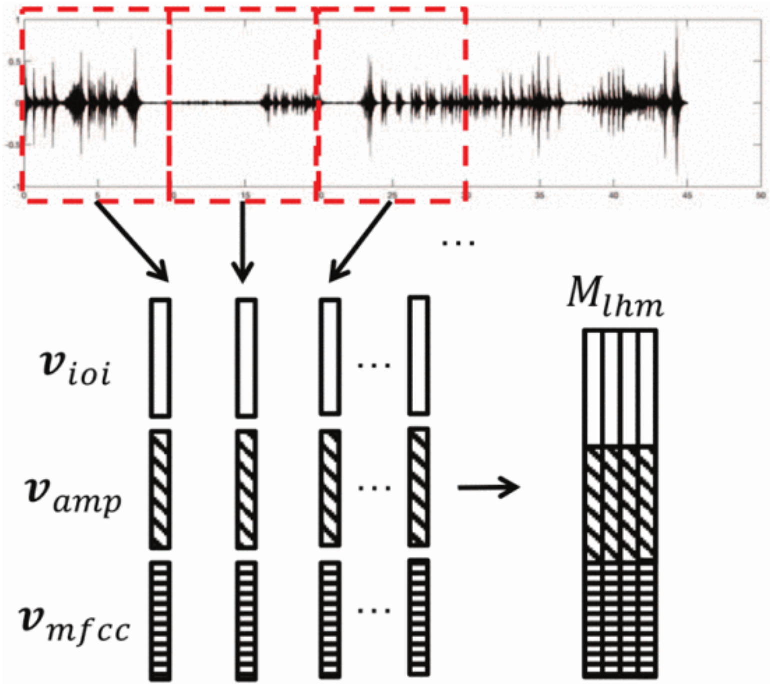


Fig. 6. The scatter view of estimated score and true score for each song

4. 타악기 연주 평가를 위한 학습된 기능

논문: https://ieeexplore.ieee.org/document/8334445





테스트에 써볼 수 있는 좋은 Midi 데이터셋

1. MAPS Database: a Piano database for multipitch estimation and automatic transcription of music: https://adasp.telecom-paris.fr/resources/2010-07-08-maps-database/
2. The MAESTRO Dataset: https://magenta.tensorflow.org/datasets/maestro