

전자식 조향 장치에서 발생하는 실차 소음 예측을 위한 다중 센서 스펙트로그램 트랜스포머

고은지¹ · 남규환² · 김상욱² · 박경환² · 김성범^{1*}¹고려대학교 산업경영공학과 / ²현대 모비스

Multi-sensor Spectrogram Transformer Network for Automobile Noise Prediction from Electric Power Steering

Eunji Koh¹ · Gue Hwan Nam² · Sang Wook Kim² · Kyung Hwan Park² · Seoung Bum Kim¹¹Department of Industrial and Management Engineering, Korea University²Hyundai Mobis

Noise level management has become an important task in the automotive industry because of the increased demand for low-noise automobiles. In particular, it is essential to reduce the automobile noises incurred by electric power steering(EPS) and an automotive steering system. Although existing methods measure the noise levels by mounting EPS on actual automobiles and detect the importance of frequency bands based on experts' judgment, they were subjective, time consuming and expensive. Therefore, an efficient method is required to predict the automobile noise levels and detect influential frequency bands based on EPS acceleration data. We propose a multi-sensor spectrogram transformer(MuST) for predicting the EPS automobile noise levels. The proposed method allows us to predict automobile noise levels by reflecting the individual characteristics of multiple sensors in EPS and detecting influential frequency bands related to noise levels. The experimental results showed that the proposed MuST performed well and detected influential frequency bands similar to the experts. We believe that the framework presented in this study can efficiently identify the automobile noise levels and help experts design quiet automobiles in the future by providing information on influential frequency bands. In addition, our method can help to handle various tasks which use data collected through multiple sensors in the automotive industry.

Keywords: Automobile Noise Prediction, Electric Power Steering, Vision Transformer Network, Feature Extraction, Important Frequency Detection

1. 서론

오늘날 저소음 차량에 대한 수요가 증가함에 따라 차량 산업에서는 실차 소음 수준이 중요한 차량 품질 평가 항목이 되었다. 실차 소음을 야기하는 부품은 전자식 파워 스티어링(electric power steering; EPS), 타이어, 모터 등이 있다(Azadi et

al., 2009). 이 중 전자식 차량 조향 장치인 EPS가 차량 주행 중 동작하며 발생하는 가속도는 실차 소음 수준에 큰 영향을 미친다(Islam and Husain, 2010). 따라서 제작된 EPS가 차량에 탑재되었을 때 유발하는 실차 소음 수준을 파악하고, 원인이 되는 주요 주파수 대역을 탐지하여 EPS의 품질을 개선하는 것이 매우 중요하다(Gieras et al., 2018).

This research was supported by BK21 FOUR.

* 연락저자 : 김성범 교수, 02841 서울특별시 성북구 안암로 145, 고려대학교 산업경영공학과, Tel : 02-3290-3397, Fax : 02-3290-4550,

E-mail : sbkim1@korea.ac.kr

2022년 1월 21일 접수; 2022년 4월 6일 수정본 접수; 2022년 4월 19일 게재 확정.

실제 차량 산업 현장에서는 EPS의 품질 평가 및 개선을 위해 EPS를 차량에 탑재하여 주행 중 발생하는 실차 소음 수준을 측정하고 전문가가 직접 중요한 주파수 대역을 탐지한다. 그러나 기존 EPS 품질 평가 방법은 고비용의 설비가 필수적이며 전문가의 경험이나 전문 지식과 같은 주관적인 요소에 의존하기 때문에 오판의 가능성이 있다. 이러한 문제를 해결하기 위해 효율적인 EPS 품질 평가 및 개선에 대한 연구가 진행되고 있다. Kim *et al.*(2011)은 자기력 고조파 저감 설계 방식을 통해 EPS 소음 저감을 시도하였으며, Yoo *et al.*(2015)는 유한요소법을 사용하여 EPS에서 발생하는 가속도로 인한 소음에 큰 영향을 미치는 원인을 탐지하였다. 그러나 선행 연구들은 EPS를 차량에 탑재하였을 때 발생하는 정확한 소음 수준 파악이 어렵고 EPS 설계 이론에 기초하는 방법론이라는 점에서 여전히 전문가의 주관적 요소에 의존한다는 한계가 있다. 이를 보완하기 위해 EPS에서 발생하는 가속도 시그널을 분석할 수 있는 모델을 도입하여 정량적인 EPS 품질 평가를 수행한다면 신뢰성 있는 EPS 품질 관리가 가능할 것이다.

최근 다양한 산업에서 모델 기반으로 센서를 통해 계측한 시그널의 패턴을 분석하고 학습하는 연구가 진행되고 있다. Dörfler *et al.*(2017)은 시그널로부터 2차원 특징 벡터를 생성하고 convolution neural network(CNN)을 적용하여 오디오 분야에 적합함을 실험적으로 보였다. Huang *et al.*(2019)는 의료 분야에서 심전도 시그널을 위한 CNN을 적용하여 심장 질환 분류를 효과적으로 수행하였다. 그러나 CNN 기반의 방법론은 시그널을 구성하는 주파수 대역의 전역적인 정보보다 지역적인 정보에 치중하여 패턴을 파악한다(He *et al.*, 2016). 이를 보완하고자 CNN을 트랜스포머(transformer)로 대체하여 시그널의 전 대역 주파수가 공유하는 패턴을 반영할 수 있는 방법론들이 제안되고 있다. Dosovitskiy *et al.*(2020)은 이미지에서 패치를 분할하는 구조를 통해 이미지의 전역적인 패턴을 학습할 수 있는 vision transformer(ViT)를 제안하였다. ViT를 통해 시그널에서 추출한 스펙트로그램(spectrogram)의 모든 주파수 대역을 고려하여 특징 벡터를 추출할 수 있다. Gong *et al.*(2021)은 시그널의 연속성을 반영하고자 ViT의 패치 분할 방식을 변경하여 audio spectrogram transformer(AST)를 제안하였다. 그러나 기존 연구들은 모두 단일 센서를 통해 수집한 시그널의 패턴을 파악하기 위한 방법론이기 때문에 다중 센서를 통해 데이터를 계측하는 차량 산업에 적용하기에는 한계가 있다. 다중 센서는 단일 센서와 달리 여러 위치의 센서를 통해 동일한 대상에 대한 시그널을 계측한다. 따라서 EPS 품질 평가를 위해 다중 센서 데이터를 적절히 활용할 수 있는 네트워크 구조가 필요하다.

다중 센서를 위한 네트워크 구조는 대표적으로 단일 모델 구조와 다중 모델 구조가 있다. 단일 모델 구조는 다중 센서를 구성하는 각 센서에서 계측된 시그널을 채널 축 방향으로 모두 결합하여 하나의 입력 정보로 간주하는 방식이다. 따라서 단일 모델 구조는 하나의 모델을 사용하기 때문에 상대적으로 적은 수의

파라미터를 갖고 다중 센서를 구성하는 센서의 개수를 쉽게 확장할 수 있다. Wang *et al.*(2017)과 Zerveas *et al.*(2021)은 시계열 데이터를 활용하는 분야에서 각각 CNN과 트랜스포머 기반의 단일 모델 구조로도 좋은 분류 성능을 보임을 실험적으로 보였다. Messner *et al.*(2018)은 폐 질환 진단을 위해 다중 센서로 계측한 시그널을 채널 축 방향으로 모두 결합하여 사용하는 CNN 기반의 단일 모델 구조를 제안하였다. 그러나 단일 모델 구조는 센서별 시그널 계측 단위의 차이가 있거나 특정 센서의 두드러지는 패턴이 존재할 경우, 모델 학습 시 개별 센서의 유의미한 특징을 모두 파악하지 못하고 일부 센서의 패턴에만 의존할 위험이 있다. 반면 다중 모델 구조는 센서마다 개별 모델을 사용하는 방식이다. 따라서 다중 센서를 구성하는 모든 센서에 독립적인 모델을 적용함으로써 모델이 센서별 고유한 패턴을 파악할 수 있다. Zhu and Sheng(2009)는 웨어러블 기기의 다중 센서를 활용하여 사람의 일상 활동 유형을 분류하고자 다중 모델 구조를 제안하였으며, Ge *et al.*(2018)은 뇌 질환 탐지를 위해 다중 센서를 구성하는 개별 센서에 독립적인 CNN을 사용하는 다중 모델 구조를 제안하였다. 그러나 다중 모델 구조는 다중 센서를 구성하는 센서의 수에 비례하여 모델의 파라미터 수 또한 증가하는 단점이 있다.

따라서 본 연구는 다중 센서로 계측한 EPS 가속도 데이터에 적합한 다중 센서 스펙트로그램 트랜스포머(multi-sensor spectrogram transformer; MuST)를 제안한다. 제안 방법론은 단일 모델 구조의 장점과 다중 모델 구조의 장점을 동시에 갖지만, 기존의 단일 모델 구조 및 다중 모델 구조와는 차이가 있다. 제안 방법론은 다중 센서를 위한 포지셔널 임베딩을 사용함으로써 센서별 데이터를 채널 축으로 결합하지 않는다는 점에서 단일 모델 구조와 차이가 있다. 또한 센서 임베딩을 통해 센서별 모델을 사용하지 않고 각 센서의 고유한 특징을 고려할 수 있다는 점에서 다중 모델 구조와 차이점이 있다. 이를 통해, 제안 방법론은 적은 파라미터만으로도 센서별 패턴과 주파수 대역별 패턴을 동시에 반영하여 기존 방법론들보다 우수한 성능으로 실차 소음 수준을 예측한다. 이 뿐만 아니라 전문가와 유사하게 실차 소음 수준에 큰 영향을 미치는 주요 주파수 대역에 대한 정보를 제공할 수 있다. 본 논문의 주요 기여점은 다음과 같다.

- EPS 실차 소음 수준 예측을 위해 다중 센서를 위한 포지셔널 임베딩과 센서 임베딩이 추가된 다중 센서 스펙트로그램 트랜스포머 네트워크를 제안한다. 따라서 제안 방법론은 다중 센서를 구성하는 개별 센서의 특징과 주파수 대역별 특징을 모두 반영할 수 있다.
- 다중 센서 스펙트로그램 트랜스포머 네트워크에서 산출된 어텐션 스코어를 통해 실차 소음 수준에 크게 영향을 미치는 주요 주파수 대역 정보를 제공한다.
- 다중 센서 스펙트로그램 트랜스포머 네트워크는 EPS 실차 소음 수준 예측 이외에도 차량 산업에서 다중 센서로 수집한 다양한 데이터에 적용 가능하다.

본 논문의 구성은 다음과 같다. 제2장에서는 데이터 처리 및 특징 추출 방법과 제안하는 방법론에 대해 설명한다. 제3장에서는 제안 방법의 예측 우수성을 보이고 제안 모델과 전문가가 탐지한 주요 주파수 대역 비교를 통해 모델의 신뢰도를 검증하였으며 제4장에서는 결론 및 추후 연구 과제와 기대 효과를 서술한다.

2. 제안 방법론

본 장에서는 연구에서 사용한 EPS 가속도 데이터를 소개하고 전처리 및 특징 추출 알고리즘을 설명한다. 이후, 다중 센서 스펙트로그램 트랜스포머 네트워크의 구조를 구체적으로 설명한다.

2.1 데이터 처리 및 특징 추출

본 연구는 단일 차종에서 취득한 EPS 가속도 데이터와 이로 인해 발생하는 실차 소음 수준 데이터를 사용하였다. 입력 데이터인 EPS 가속도 데이터는 총 6개의 센서를 사용하여 EPS의 두 가지 위치인 ball nut assembly(BNA)와 motor에서 각각 X축, Y축, Z축에 대해 12,800Hz의 샘플링 비율(sampling rate)을 갖도록 계속하였다. 하나의 관측치는 차량의 핸들 회전 중 0.25초 동안 계속되는 가속도이다. 따라서 하나의 관측치에는 6개의 센서 데이터가 포함되며, 개별 센서 데이터는 3,200개의 가속도 계측 값을 갖는다. 출력 데이터인 실차 소음 수준 데이터는 핸들이 회전하는 0.25초 동안 0.1초 단위로 측정된 실차 소음 수준의 평균 값이다.

EPS 가속도는 시간에 따른 진폭 크기 변화를 나타내는 시간 영역에서 계속된다. 그러나 시간 영역은 가속도를 구성하는 각 주파수 대역의 성분 변화 정보를 제공할 수 없다는 점에서 한계가 있다. 이를 해결하기 위해 가속도를 구성하는 주파수 대역별 성분 변화를 나타내는 주파수 영역으로 표현하였다. 따라서 본 연구는 단시간 푸리에 변환 알고리즘을 적용하여 EPS 가속도를 주파수 영역에서 표현하는 스펙트로그램을 추출하였다. <Figure 1>은 시간 영역과 주파수 영역으로 표현된 가속도를 나타낸다.

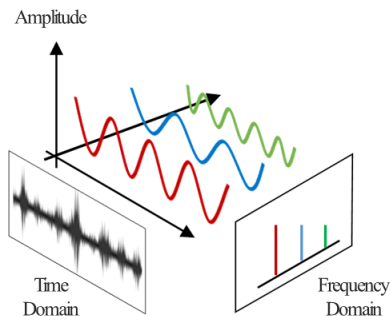


Figure 1. Time and Frequency Domains of Acceleration Signals

단시간 푸리에 변환 $F(n,w)$ 은 식 (1)과 같이 표현할 수 있다. 이 때, n 은 이동 구간, $w(t)$ 는 윈도우 함수, $x(t)$ 는 변환 전 신호 값, t 는 시간을 의미한다. 단시간 푸리에 변환에 사용된 윈도우 함수는 식 (2)와 같다. 단시간 푸리에 변환을 거친 후에 데시벨(decibel) 스케일로 변경하여 스펙트로그램을 추출한다. 스펙트로그램은 시간에 따른 주파수 대역별 성분 변화를 표현하며 X축인 시간 축과 Y축인 주파수 축으로 구성된다.

$$F(n,w) = \sum_{t=-\infty}^{\infty} x(t)w(t-n)e^{-j\omega t} \quad (1)$$

$$w(t) = 0.54 - 0.46\cos\left(\frac{2\pi t}{L}\right) \quad (2)$$

본 연구에서는 n 을 162, 윈도우 길이를 640으로 설정하여 스펙트로그램을 추출하였다. 윈도우 길이를 640으로 설정함으로써 20Hz 단위로 주파수 성분 정보를 갖는 스펙트로그램을 추출할 수 있다. 이후 전문가의 의견에 따라 3,200Hz 이상의 주파수 대역은 실차 소음 수준과 무관한 노이즈로 구성된 대역으로 판단하여 제거하였다. 최종적으로 시간 축으로 16, 주파수 축으로 160의 크기를 갖는 스펙트로그램을 생성하였다. <Figure 2>는 6개 센서로 계측한 하나의 EPS 가속도 관측치에 대하여 추출된 스펙트로그램을 시각화한 그림이다.

2.2 다중 센서 스펙트로그램 트랜스포머 네트워크 구조

본 연구에서는 다중 센서를 구성하는 개별 센서별 특징과 센서 간 공유하는 주파수 대역별 특징을 모두 학습할 수 있는 다중 센서 스펙트로그램 트랜스포머 네트워크를 제안한다. 제안 방법론은 기본적으로 ViT 구조를 따르면서 다중 센서에 적합하도록 패치 임베딩(patch embedding) 단계를 변형한다. 본 연구에서 제안한 다중 센서 스펙트로그램 트랜스포머는 패치 임베딩과 트랜스포머 인코더(encoder) 및 소음 예측기(predictor)로 구성되며 전반적인 학습 방식은 <Figure 3>에서 보여주고 있다. 네트워크의 각 단계를 아래에서 구체적으로 설명한다.

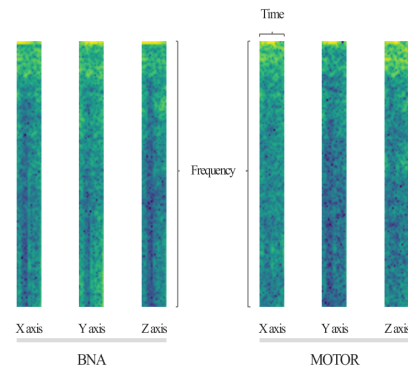


Figure 2. Spectrograms of Sample

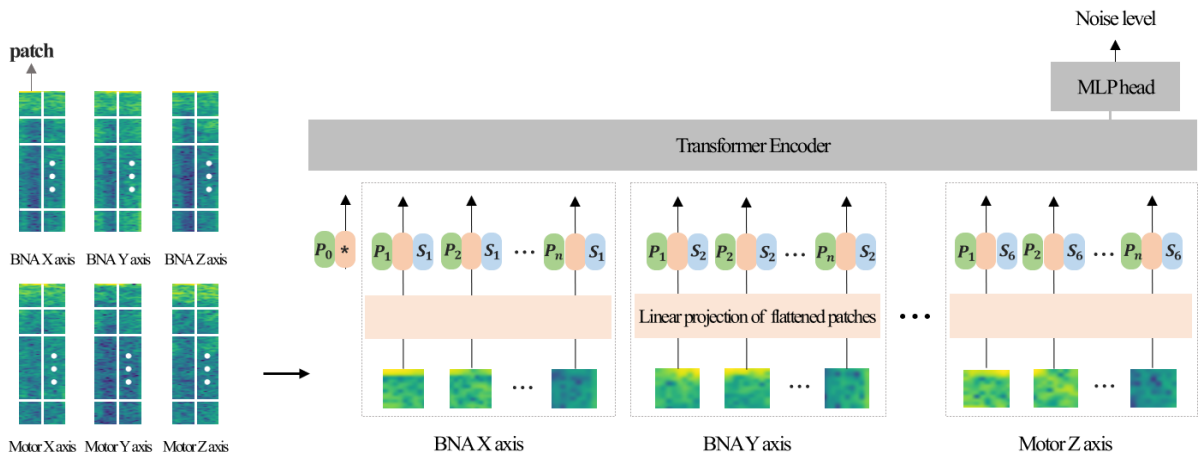
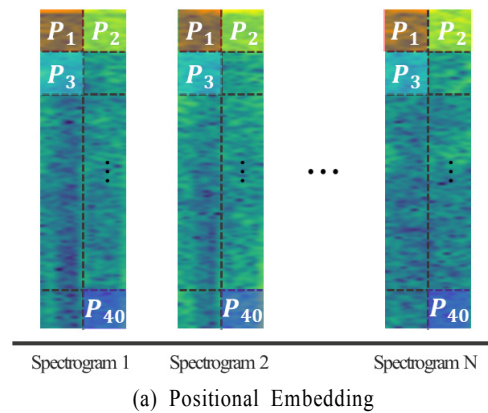


Figure 3. Architecture of Multi-Sensor Spectrogram Transformer (MuST)

패치 임베딩: 스펙트로그램 $X \in R^{H \times W \times C}$ 를 균일한 크기의 2차원 패치 $X_p \in R^{N \times (P^2 \cdot C)}$ 로 분할하여 순차적으로 나열한다. 이때, (H,W)는 원본 스펙트로그램 해상도(resolution), C는 채널 수, (P,P)는 패치의 해상도를 의미한다. 단일 스펙트로그램에서 생성된 패치의 수는 $N = HW/P^2$ 으로 표현할 수 있다. <Figure 4>는 스펙트로그램에서 분할된 패치를 시각화한 것이다. 다음으로 나열한 패치들을 학습 가능한 linear projection을 통해 D차원의 특징 벡터(feature vector)로 각각 매핑(mapping)한다. 단일 스펙트로그램에 대한 linear projection은 식 (3)으로 표현되며 projection의 결과를 패치 임베딩이라고 정의한다. 모든 스펙트로그램에 대해 패치 임베딩을 수행하여 얻은 특징 벡터를 결합한다.

패치 임베딩에 학습 가능한 값인 포지셔널 임베딩(positional embedding)과 센서 임베딩(sensor embedding)을 더함으로써 모델이 센서별 고유한 패턴과 센서 간 공유하는 패턴을 모두 반영할 수 있도록 한다. <Figure 5>는 패치 임베딩과 센서 임베딩을 표현한다. 이때, P_i 는 원본 스펙트로그램 내 i 번째 패치의 포지셔널 임베딩, S_j 는 j 번째 센서에 대한 센서 임베딩을 의미

한다. 포지셔널 임베딩은 다른 센서에서 추출된 스펙트로그램 임베딩도 동일한 주파수 대역 및 시간 대의 패치 간 공유하는 패턴이 있을 것으로 가정하고 부여한다. 포지셔널 임베딩은 원본 스펙트로그램 내 패치의 위치 정보를 제공하기 위한 것이기 때문이다. 따라서 원본 스펙트로그램에서 패치 위치에 따라 다른 포지셔널 임베딩이 부여되며 센서의 종류와 관계없이 모든 스펙트로그램에서 포지셔널 임베딩을 공유한다.



(a) Positional Embedding

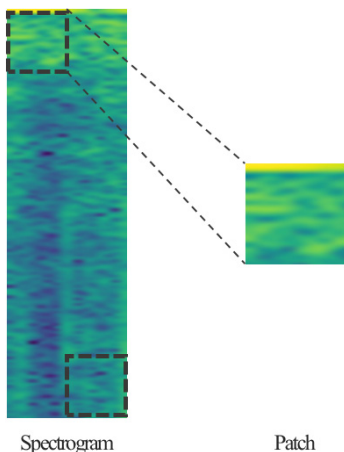
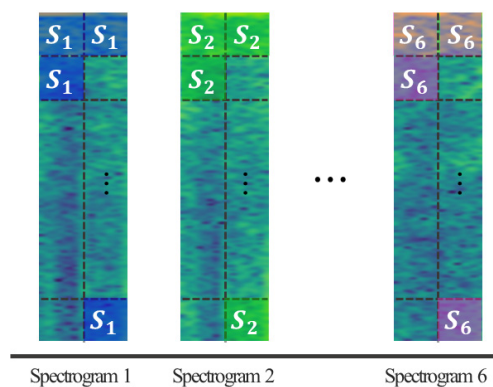


Figure 4. A Patch Split from Spectrogram



(b) Sensor Embedding

Figure 5. Positional Embedding and Sensor Embedding of MuST

다음으로 센서 임베딩은 개별 센서에서 추출된 스펙트로그램이 고유한 패턴을 갖을 것으로 가정하고 부여한다. 센서 임베딩은 패치가 분할된 스펙트로그램의 센서 종류 정보를 제공하는 것이 목적이기 때문이다. 따라서 동일한 스펙트로그램에서 분할된 패치들에 동일한 센서 임베딩이 부여되며 서로 다른 스펙트로그램에서 분할된 패치들에는 다른 센서 임베딩이 부여된다.

트랜스포머 인코더 및 소음 예측기: <Figure 6>는 트랜스포머 인코더의 자세한 구조를 나타낸다. 트랜스포머 인코더는 layer normalization(LM)(Ba *et al.*, 2016)과 multi-head self-attention(MSA) 블록 및 multi-layer perceptron(MLP) 블록으로 구성된다. MSA는 식 (4)로 나타낼 수 있으며 비선형 함수인 Gaussian error linear unit(GELU)(Hendrycks *et al.*, 2016)을 사용하는 두 개의 layer로 구성되는 MLP는 식 (5)으로 표현된다. 이 과정에서 각 패치의 중요도를 정량적으로 표현한 어텐션 스코어를 통해 제안 방법론이 상대적으로 큰 가중치를 두며 집중한 패치를 파악할 수 있다. 본 연구에서는 모델이 실차 소음 수준에 큰 영향을 미치는 주파수 대역에 크게 집중할 것으로 가정하고 어텐션 스코어를 통해 주요 주파수 대역 정보를 제공하고자 한다. 이후 최종적으로 트랜스포머 인코더를 통해 파악한 스펙트로그램들의 패턴을 바탕으로 MLP로 구성된 소음 예측기를 거쳐 실차 소음 수준을 예측한다.

$$Z_0 = [X_{dass}; X_p^1 E; X_p^2 E; \dots; X_p^n E] + E_{pos} + E_{sensor}, \quad (3)$$

$$E \in R^{(P^2 \cdot G) \times D}, E_{pos} \in R^{(N+1) \times D}, E_{sensor} \in R^{(N+1) \times D}$$

$$z'_l = MSA(LN(z_{l-1})) + z_{l-1}, l = 1, \dots, L \quad (4)$$

$$z_l = MLP(LN(z'_l)) + z'_l, l = 1, \dots, L \quad (5)$$

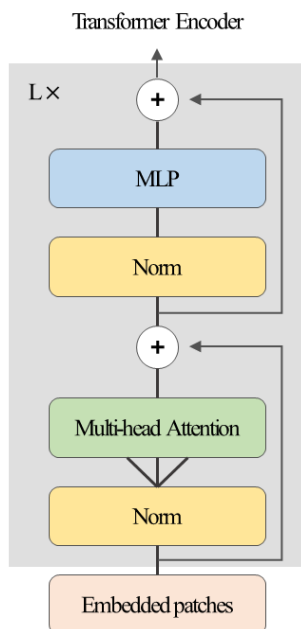


Figure 6. Architecture of Transformer Encoder

3. 실험 결과

본 장은 제안 방법론 성능 평가 실험과 관련된 하이퍼파라미터 설정, 성능 평가 지표, 제안 방법론의 실험 결과를 포함한다. 제안 방법론이 단일 모델 구조임에도 개별 센서의 패턴을 반영할 수 있음을 검증하고자 센서 임베딩 적용 여부에 따른 두 가지 네트워크에 대해 실차 소음 수준 예측 성능을 평가하였다. 아울러 전문가 의견과 어텐션 스코어의 비교를 통해 제안 방법론이 신뢰성 있게 중요 주파수 대역 정보를 제공할 수 있음을 검증하였다.

3.1 실험 설정

본 연구에서는 MuST의 하이퍼파라미터로 패치 높이와 넓이를 8로 설정하였다. 패치 높이를 8로 설정함으로써 20Hz 단위의 주파수 대역 성분으로 구성된 스펙트로그램에서 생성된 각 패치는 160Hz 단위로 주파수 대역의 특징을 파악한다. 학습에 필요한 손실 함수는 예측 문제에 주로 사용되는 mean square error(MSE) loss, optimizer 함수는 AdamW(Loshchilov *et al.*, 2017)를 사용하였으며 학습률(learning rate)은 0.001로 설정하였다. 학습 시 사용한 batch size는 64, epoch는 50이다. 트랜스포머 인코더는 10개의 층(layer)로 이루어져 있으며 각 층은 8개의 head를 갖는 MSA 블록과 MLP 블록으로 구성된다.

제안 방법론을 평가하기 위해 CNN, ViT, AST를 기반으로 하는 단일 모델 구조의 네트워크와 다중 모델 구조의 네트워크를 비교 방법론으로 사용하였다. CNN 모델은 ResNet18과 ResNet34를 사용하였다(He *et al.*, 2016). 검증 데이터를 사용하여 최적 하이퍼파라미터를 선정하였고, 테스트 데이터에 대한 예측 성능을 측정하였다. 전체 관측치 수는 11,359개로 학습 데이터는 6,317개, 검증 데이터는 724개, 테스트 데이터는 4,117개를 사용하였다. 또한, 제안 방법론과 비교 방법론의 성능이 유의한 차이가 있는지 확인하기 위해 서로 다른 다섯 가지 데이터 세트를 구성하여 반복 실험하였다.

3.2 실차 소음 수준 예측 성능

제안 방법론과 비교 방법론의 실차 소음 수준 예측 성능을 정량적으로 평가하고자 결정 계수(R^2)와 평균 절대 오차(mean absolute error; MAE)를 사용하였다. 결정 계수는 식 (6), 평균 절대 오차는 식 (7)로 표현할 수 있다. 이때, 평가 지표들은 실제 소음 수준 Y_k , 실제 소음 수준의 평균 \bar{Y} , 예측 값 \hat{Y}_k , 전체 관측치의 수 n 으로 구성된다. 결정 계수는 출력 데이터의 변동에 대한 입력 데이터의 설명력을 나타내는 지표이며, 출력 데이터의 평균을 예측 값으로 사용했을 때 대비 모델을 통한 예측 값의 설명력을 비교한다. 평균 절대 오차는 실제 소음 수준과 모델을 통한 예측 값의 차이에 절대값을 취하여 평균을 낸 값이다.

$$R^2 = 1 - \frac{\sum_{k=1}^n (Y_k - \hat{Y}_k)^2}{\sum_{k=1}^n (Y_k - \bar{Y})^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{k=1}^n |Y_k - \hat{Y}_k| \quad (7)$$

<Table 1>은 제안 방법론과 비교 방법론의 예측 성능을 나타낸다. 제안 방법론은 EPS 소음 수준 예측에 대한 결정 계수가 0.73, 평균 절대 오차가 0.94로 비교 방법론들과 대비해 가장 뛰어난 예측 성능을 보였다. 특히 차량 산업 전문가들이 차량 내부 소음 평가 시, 평균 절대 오차가 2dB 이내여야 의미가 있다는 의견에 비추어 볼 때 제안 방법론이 우수한 성능을 내는 것으로 평가할 수 있다. <Figure 7>은 실제 소음 값에 따른 제안 방법론의 예측 결과를 나타낸다. <Figure 7>에 제시된 산점도의 X축과 Y축은 각각 실제 소음 값과 제안 방법론의 예측 소음 값을 의미한다. 산점도에 점선으로 표현된 Y = X 그래프와의 비교를 통해 제안 방법론은 대체로 실제 소음 값과 유사하게 소음을 예측하는 것을 확인할 수 있다. <Table 1>의 실험 결과에 따르면 사용한 모델과 관계없이 다중 모델 구조 비교 방법론에서 상대적으로 낮은 성능을 보인다. 따라서 다중 모델 구조는 주로 센서별 특징에 주목하여 실차 소음을 예측하기 때문에 주파수 대역의 특징을 잘 파악하지 못하는 것으로 해석할 수 있다. 반면 주파수 대역별 특징을 주로 반영하는 단일 모델 구조 비교 방법론은 다중 모델 구조 비교 방법론 대비 상대적으로 높은 성능을 갖는다. 그러나 제안 방법론에는 미치지 못하는 성능이다. 아울러 제안 방법론은 비교 방법론들 대비 필요한 파라미터 개수가 적다는 장점이 있으며 이를 통해 단일 모델 구조임에도 불구하고 EPS 실차 소음 수준을 높은 성능으로 예측함을 확인하였다.

Table 1. Comparison Results of Models. Boldface Values Represent the Best Performance and Standard Deviations are Included Next to the MAE and R^2

Models	Network Structure	MAE	R-square	Parameter
ResNet18	Single model	1.30 ± 0.26	0.46 ± 0.33	11,175,681
	Multi model	1.35 ± 0.18	0.48 ± 0.13	67,039,367
ResNet34	Single model	1.67 ± 0.29	0.21 ± 0.24	1,175,681
	Multi model	1.52 ± 0.16	0.24 ± 0.18	127,688,327
ViT	Single model	1.04 ± 0.14	0.66 ± 0.11	524,161
	Multi model	1.38 ± 0.33	0.43 ± 0.18	3,024,647
AST	Single model	1.06 ± 0.15	0.64 ± 0.14	524,673
	Multi model	1.44 ± 0.27	0.31 ± 0.18	3,027,719
MuST (Proposed Method)		0.94 ± 0.03	0.73 ± 0.01	501,121

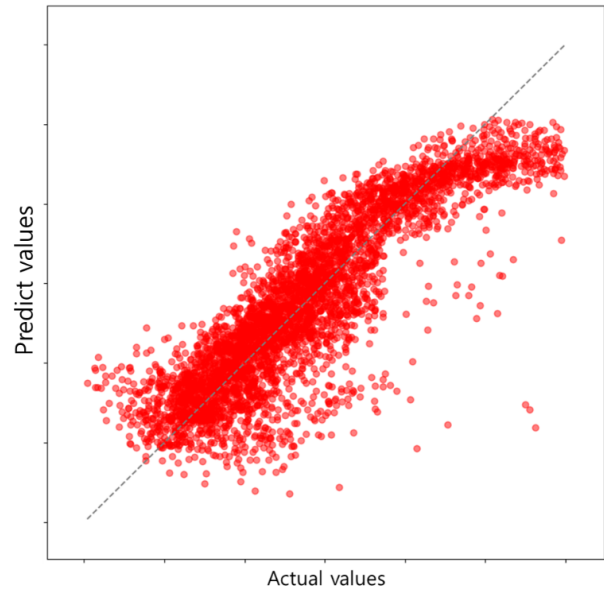


Figure 7. Scatter Plot for Results of MuST

Table 2. Comparison among Models in Terms of Sensor Embedding. Boldface Values Represent the Best Performance and Standard Deviations are Included Next to the MAE and R^2

Models	MAE		Parameter
MuST with only positional embedding	0.98 ± 0.06	0.69 ± 0.04	501,121
MuST with positional and sensor embedding	0.94 ± 0.03	0.73 ± 0.01	501,121

<Table 2>에서는 센서 임베딩 여부에 따른 제안 방법론의 소음 예측 성능을 비교하였다. 센서 임베딩을 사용하지 않을 때의 예측 성능은 결정 계수 0.69로 트랜스포머 기반으로 하는 단일 모델 구조 비교 방법론의 성능과 차이가 적음을 알 수 있다. 그러나 센서 임베딩을 사용하는 경우, 결정 계수 0.73으로 센서 임베딩을 사용하지 않는 경우보다 큰 성능 향상을 보인다. 따라서 제안 방법론의 센서 임베딩이 효과적으로 센서별 특징을 학습함을 확인하였다.

3.3 중요 주파수 영역 추출 과정

EPS에서 발생하는 가속도로 인한 실차 소음 수준을 통해 EPS 품질 관리를 하기 위해서는 소음 수준에 큰 영향을 미치는 주요 주파수 대역 정보를 파악할 수 있어야 한다. 제안 방법론은 어텐션 메커니즘에서 산출된 어텐션 스코어를 통해 EPS 실차 소음 수준에 큰 영향을 미치는 주요 주파수 대역을 파악할 수 있었다. <Figure 8>는 센서별 주파수 대역의 중요도를 히트맵으로 나타내었으며 어텐션 스코어가 높게 산출된 주파수 대역일수록 붉은 색으로 표시하였다. 히트맵 결과를 통해 센

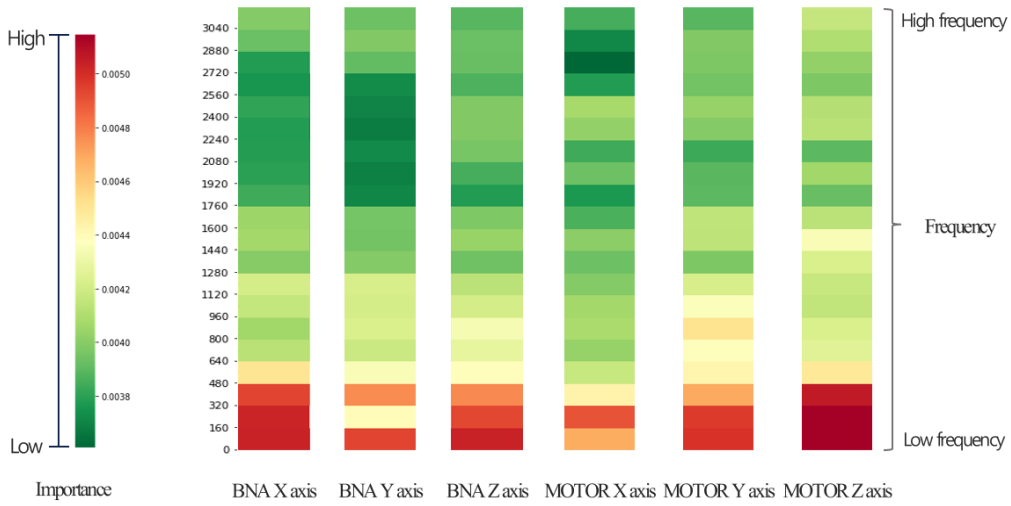


Figure 8. Heatmaps for Representing the Importance of Frequency Bands in each Sensor

서 종류와 관계 없이 제안 방법론이 산출한 주요 주파수 대역은 480Hz 이하임을 알 수 있다.

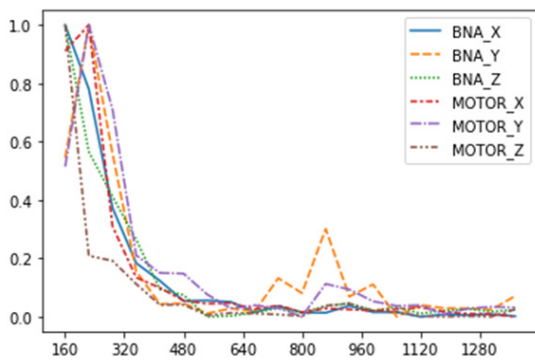
제안 방법론이 제공하는 주요 주파수 대역 정보의 신뢰성을 입증하기 위해 실제 차량 산업 현장에서 활용하는 주요 주파수 대역 탐색 방법과 비교하였다. 차량 산업 현장에서는 noise transfer function(NTF)을 통해 주요 주파수 대역을 탐지하고 있다. NTF를 산출하기 위해 우선 차량의 EPS 부근에 인위적인 타격을 주어 소음을 발생시킨다. 다음으로 발생한 소음을 통해 주파수에 따른 응답 정보를 파악하여 응답 정도가 큰 주파수를 중요한 주파수로 산출한다. NTF 결과에 따라 차량 산업 전문가들은 통상적으로 300Hz 이내에서 주요 주파수 대역이 나타난다고 판단하고 있다.

<Figure 9>는 NTF 방식을 통한 주요 주파수 대역 탐지 방법과 제안 방법론이 산출한 주파수 대역별 중요도를 그래프로 보여주고 있다. <Figure 9>에 따르면 NTF 결과와 제안 방법론을 통해 산출한 주요 주파수 대역이 모두 저주파수 대역에서 나타남을 알 수 있다. 따라서 제안 방법론은 NTF를 바탕으로 주요 주파수를 탐색하는 전문가와 유사한 양상으로 각 주파수 대역의 중요도를 파악하는 것으로 해석할 수 있다.

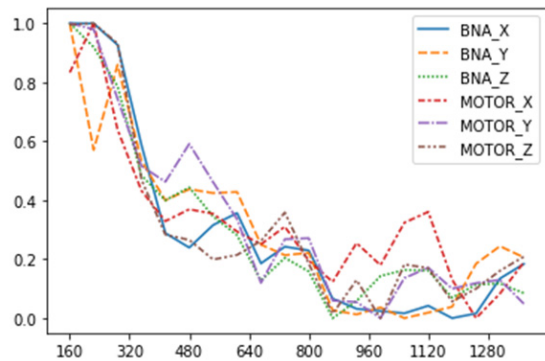
Table 3. Comparison of Importance Ranking by Frequency Band Calculated from MuST and NTF

Rank	MuST	NTF
1	0 - 160Hz	0 - 160Hz
2	160 - 320Hz	160 - 320Hz
3	320 - 480Hz	320 - 480Hz
4	480 - 640Hz	480 - 640Hz
5	800 - 960Hz	640 - 800Hz
6	640 - 800Hz	1760 - 1920Hz
7	960 - 1120Hz	800 - 960Hz
8	1120 - 1280Hz	1920 - 2080Hz
9	1440 - 1600Hz	1440 - 1600Hz
10	1600 - 1760Hz	2080 - 2240Hz

<Table 3>은 NTF의 결과와 제안 방법론의 어텐션 스코어에 따라 주파수 대역별 중요도를 순위에 따라 제시한다. <Table 3>에 따르면 모든 센서의 주파수 대역별 평균 중요도에 대해 상위 네 개의 주파수 대역이 일치하였다. 상위 네 개



(a) Experts



(b) Model

Figure 9. Line Plot for Representing the Importance of Frequency Bands in each Sensor

주파수 대역은 0Hz에서 640Hz로 저주파수 대역이다. 또한, 상위 10개 주파수 대역을 비교하였을 때 대체로 유사한 양상을 보였다. 따라서 다중 센서 스펙트로그램 트랜스포머 네트워크는 전문가와 유사하게 저주파수 대역에 집중하여 EPS 실차 소음 수준을 추론하고 있음을 알 수 있었다. 아울러 제안 방법론은 실제 차량 산업에서 EPS 실차 소음 수준을 예측하는 것뿐만 아니라 전문가의 의견과 유사하게 실차 소음 수준에 큰 영향을 미치는 주요 주파수 대역 정보를 제공할 수 있음을 보였다.

4. 결론

본 연구는 EPS 가속도 데이터를 사용하여 실차 소음 수준을 예측하는 다중 센서 스펙트로그램 트랜스포머 네트워크를 제안한다. EPS 가속도 데이터는 여섯 개의 다중 센서를 통해 수집되었다. 가속도 시그널을 주파수 영역으로 표현하고자 단시간 푸리에 변환 알고리즘을 적용하여 센서별 스펙트로그램을 추출하였다. 제안 방법론은 다중 스펙트로그램의 센서별 특징에 주목하는 센서 임베딩과 스펙트로그램 간 공유하는 주파수 대역별 특징을 주로 반영하는 포지셔널 임베딩을 포함한다. 기존 ViT 구조와 달리 다중 센서를 위한 포지셔널 임베딩과 센서 임베딩을 사용함으로써, 제안 방법은 비교 방법들보다 적은 파라미터만을 사용함에도 불구하고 우수한 실차 소음 예측 성능을 보이는 것을 확인하였다. 아울러 제안 방법론은 실차 소음 수준에 큰 영향을 미치는 주요 주파수 대역 정보를 제공할 수 있다. 실제로 NTF 방법을 통해 전문가가 탐지한 주요 주파수 대역과 제안 방법론의 어텐션 스코어를 통해 산출한 주요 주파수 대역이 유사하게 나타남을 보였다. 본 연구는 다중 센서를 통해 가속도를 계측하는 차량 산업에 적합한 새로운 네트워크 모델이라는 점에서 의의가 있다. EPS의 소음 수준 예측과 주요 주파수 대역 정보를 제공함으로써 차량 산업 현장에서 EPS 품질 개선에 기여할 수 있을 것으로 판단한다. 이뿐만 아니라 차량 산업에서 EPS 가속도 데이터 이외의 다중 센서 데이터를 이용한 소음 예측 문제에 다방면으로 활용할 수 있을 것으로 기대한다.

본 연구에서 생성한 스펙트로그램 패치의 크기는 160Hz 단위의 주파수 대역에서 유의미한 패턴이 나타난다는 전문가 의견에 따라 높이와 넓이를 각각 8로 설정하였다. 하지만 160Hz 단위 이상의 넓은 주파수 대역에서 유의미한 패턴이 나타나는 경우 모델 성능이 저하될 수 있다는 한계가 있다. 이를 보완하기 위해 향후에는 패치를 반복적으로 생성하며 패치 생성 위치를 지속적으로 업데이트할 수 있는 유의미한 패턴 중심의 패치 생성 방식(Yue et al., 2021)을 적용할 예정이다. 패치 생성 위치가 업데이트됨에 따라 점차 유의미한 패턴이 나타나는 주파수 대역 위주로 패치를 생성하게 되며 이는 패치 생성 시 유의미한 패턴의 구조가 분할되지 않는다는 이점을 갖는다.

참고문헌

- Azadi, S., Azadi, M., and Zahedi, F. (2009), NVH analysis and improvement of a vehicle body structure using DOE method, *Journal of Mechanical Science and Technology*, **23**(11), 2980-2989.
- Ba, J. L., Kiro, J. R., and Hinton, G. E. (2016), Layer normalization, arXiv preprint arXiv:1607.06450.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Housby, N. (2020), An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929.
- Dörfler, M., Bammer, R., and Grill, T. (2017, July), Inside the spectrogram: Convolutional Neural Networks in audio processing, In *2017 International Conference on Sampling Theory and Applications (SampTA)*, IEEE, 152-155.
- Ge, C., Gu, I. Y. H., Jakola, A. S., and Yang, J. (2018, July), Deep learning and multi-sensor fusion for glioma classification using multistream 2D convolutional networks, In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 5894-5897.
- Gieras, J. F., Wang, C., and Lai, J. C. (2018), *Noise of polyphase electric motors*, CRC press.
- Gong, Y., Chung, Y. A., and Glass, J. (2021), AST: Audio Spectrogram Transformer, arXiv preprint arXiv:2104.01778.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016), Deep residual learning for image recognition, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.
- Hendrycks, D. and Gimpel, K. (2016), Gaussian error linear units (gelus), arXiv preprint arXiv:1606.08415.
- Huang, J., Chen, B., Yao, B., and He, W. (2019), ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network, *IEEE Access*, **7**, 92871-92880.
- Islam, R. and Husain, I. (2010), Analytical model for predicting noise and vibration in permanent-magnet synchronous motors, *IEEE Transactions on Industry Applications*, **46**(6), 2346-2354.
- Kim, J., Kim, D., Hong, J., and Lee, G. (2011), Improvement of Acoustic Noise Characteristics by Reducing Electromagnetic Force Harmonics of PMSM for EPS, *Trans. KSAE*, 1106-1110.
- Loshchilov, I. and Hutter, F. (2017), Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101.
- Messner, E., Fediuk, M., Swatek, P., Scheidl, S., Smolle-Jüttner, F. M., Olschewski, H., and Pernkopf, F. (2020), Multi-channel lung sound classification with convolutional recurrent neural networks, *Computers in Biology and Medicine*, **122**, 103831.
- Wang, Z., Yan, W., and Oates, T. (2017, May), Time series classification from scratch with deep neural networks: A strong baseline, In *2017 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 1578-1585.
- Yoo, B., Niguchi, N., Hirata, K., and Zaini, A. (2015, May), Identification of acoustical noise source for electric power steering motor using finite element method, In *2015 IEEE International Magnetics Conference (INTERMAG)*, IEEE, 1-1.
- Yue, X., Sun, S., Kuang, Z., Wei, M., Torr, P. H., Zhang, W., and Lin, D. (2021), Vision transformer with progressive sampling, In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 387-396.
- Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., and Eickhoff,

C. (2021, August), A transformer-based framework for multivariate time series representation learning, In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2114-2124.

Zhu, C. and Sheng, W. (2009, May), Human daily activity recognition in robot-assisted living using multi-sensor fusion. In *2009 IEEE International Conference on Robotics and Automation*, IEEE, 2154-2159.

저자소개

고은지 : 성신여자대학교 경영학과에서 2021년 학사학위를 취득하고 고려대학교 산업경영공학과에서 석사과정에 재학 중이다. 연구분야는 Machine Learning, Deep Learning for Multi-variate Time Series Data이다.

남규환 : 한양대학교 기계공학부에서 2017년 학사학위를 취득하고 현대모비스 해석연구센터 NVH응용기술센터에서 연구원으로 재직 중이다. 연구분야는 차량 조향장치 NVH 평가, NVH 데이터 분석이다.

김상욱 : 고려대학교 기계공학과에서 2017년 학사학위를 취득하고 현대모비스 해석연구센터 NVH응용기술센터에서 연구원으로 재직 중이다. 연구분야는 차량 조향장치 NVH 평가, NVH 데이터 분석이다.

박경환 : 셀리더는 고려대학교 기계공학과에서 1993년 학사, 1995년 석사학위를 취득하고 1995년부터 현대모비스 연구소에서 차량 소음진동 분야 연구원으로 활동 중이다.

김성범 : 고려대학교 산업경영공학부 교수로 2009년부터 재직하고 있으며, 인공지능공학연구소 소장 및 기업산학협력센터 센터장을 역임했다. 미국 University of Texas at Arlington 산업공학과에서 교수를 역임하였으며, 한양대학교 산업공학과에서 학사학위를 미국 Georgia Institute of Technology에서 산업공학 석사 및 박사학위를 취득하였다. 인공지능, 머신러닝, 최적화 방법론을 개발하고 이를 다양한 공학, 자연과학, 사회과학 분야에 응용하는 연구를 수행하고 있다.