Predicting Ground Reaction Forces with EMG Data

Benjamin Randoing
Department of Mechanical Engineering
Stanford University
bar39@stanford.edu

Quincy Delp
Department of Mechanical Engineering
Stanford University
qdelp@stanford.edu

Cooper Shea
Department of Mechanical Engineering
Stanford University
coopshea@stanford.edu

1 Abstract

The development of musculoskeletal models allows us to diagnose and treat pathological movement patterns as well as evaluate and increase athletic performance. High quality models are typically the result of over-constrained optimization problems informed by many measurements. Electromyography (EMG) measures muscle activation, internal measurement units (IMUs) measure movement, force plates measure ground reaction forces (GRFs), and motion-capture systems measure kinematics. GRFs are foundational in many biomechanical analyses but their direct measurement typically limits the scope of a study because force plates restrict the field through which a subject can move. In contrast, EMG measurements can be gathered remotely with untethered sensors attached directly to the subject. Consequently, indirect measurement of GRFs, predicted via a model, would allow for a wider variety of experiments. Here we demonstrate that machine learning architectures are capable of modeling the relationship between EMG data and GRFs and can compute GRF estimations in real time. We evaluate the performance of four neural networks trained on paired EMG and GRF data captured as three subjects walked at a steady pace. The models can independently predict forces for each limb and identify gait cycle features like heel-contact and push-off. While this study only examines subjects with typical gait patterns performing a single activity, the success of the models motivates more analysis of how we can indirectly measure GRFs. More broadly, results like this validate tactics to design studies with a wider variety of motions, not limited to the area covered by a force plate.

2 Introduction

2.1 Problem

Ground reaction forces are a powerful diagnostic tool for medical professionals. Notably, ground reaction forces are commonly used to asses lower limb support in stroke patients [4]. In addition to their impact in clinics, GRFs support the practice of inverse dynamics for biomechanists performing simulations for research studies. After motion capture data collection, GRFs allow for a more resolute result from inverse dynamics. Despite the impact GRFs have in many settings, common practices of obtaining GRFs involve instrumented force plates that pose significant constraints on many researchers and medical teams as a result of their size and weight. The inability to measure GRFs for long durations and in everyday contexts limits the potential insight that may be obtained from GRFs. Therefore, solving the need for a portable alternative to measuring GRFs would support the development of studies and practices that monitor human movement during an individuals every routine.

2.2 Previous Works

Several previous works have focused on predicting ground reaction forces (GRFs) using different approaches and data sources. One common approach has been to utilize accelerometer data from an inertial measurement unit (IMU) to estimate GRFs. In one study, a single accelerometer was placed on the sacrum of subjects to obtain an IMU data stream. This data was then transformed into GRFs by applying a rotation matrix to align the data with the ground coordinate axis. The resulting acceleration profile in the vertical axis was multiplied to derive the force in the vertical direction [2].

In addition to IMU data, previous attempts to generate GRFs have explored the use of algorithmically complex neural network architectures, such as echo state networks. These networks have been employed to predict joint kinematics and GRFs based on input data streams from surface electromyography (EMG) [1]. By leveraging the cyclical nature of biomechanical data, including EMG and gait kinematics signals, these models have shown promise in predicting GRFs. However, there is a growing need to develop models with greater computational simplicity to enable real-time prediction of GRFs for applications in clinical and research settings.

Surface EMG (sEMG) has emerged as a commonly used form-factor for collecting EMG data. Compared to implanted EMGs, surface EMGs offer increased safety and comfort [3]. However, data obtained from surface-level EMGs presents certain challenges that require additional post-processing. Signal artifacts arising from soft tissue movement and placement errors necessitate rectifying, filtering, and normalizing the data stream [6]. Moreover, it is common practice to conduct maximal effort trials to establish an activation baseline and develop more accurate models. Despite these challenges, surface EMGs have been successfully employed in studies involving spinal cord injury and step-down activation in patients with osteoarthritis [5, 1].

2.3 Model Architectures

2.3.1 Simple Linear and Fully Connected Models

Two regression models, namely direct linear regression and a fully connected multi-layer perceptron (MLP), were employed to explore the relationship between surface electromyography (EMG) signals and ground reaction force. The intuitive expectation of a linear association between these variables stems from the observation that during the stance phase of the gait cycle when the leg is in contact with the ground, muscle activations in the leg are higher compared to the swing phase. The shallow fully connected neural network was chosen to strike a balance between simplicity and the ability to capture subtle nonlinearities in the ground reaction force profile across different gait patterns. Moreover, the adoption of a shallow architecture maintains a high degree of interpretability, setting it apart from other complex models commonly used in the field.

2.3.2 WaveNet

WaveNet is a MLP architecture that uses dilated convolutions to predict the next element in a time-series data stream[9]. It was developed by Google for generating audio outputs based on the previous context of an input audio signal. For example, generating musical audio with a WaveNet architecture empowers a neural network to generate an audio output based on the previous 1024 audio amplitudes. The WaveNet architecture is depicted in Figure 2*** Audio and EMG data streams are both time-series data that contain significant high-frequency content. The ability to use a receptive field, conditioning each GRF prediction on several timesteps of EMG information seemed powerful and, unlike the audio-to-audio model trained in the original paper, we were able to implement a lookahead by incorporating a short (100ms) delay. While we were inspired by the original architecture, we made modifications to adapt it to our use case. We used 2d convolutions and a receptive field of 16 with a lookahead of 7.

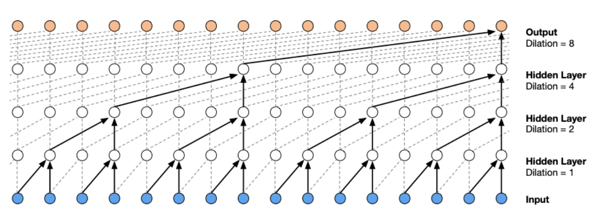


Figure 1: WaveNet Dilated Convolution Architecture [8]

2.3.3 Echo State Network

In preparing the literature review in response to the hopeless confusion of trying to understand the WaveNet model, Cooper came across the Echo State Network used in [1]. This paper is the closest to our final goal - so Cooper took a shot at implementing this approach. In the same paper, it is found that Recurrent and Convolutional Neural Nets (RNNs, CNNs) very slightly outperformed the ESN (.94 and .93 vs. .92) in finding GRFs, but under-performed on finding joint angles. In addition, ESN, CNN, and RNN all vastly outperform simpler, traditional machine learning approaches (mentioned in 3.0) at both joint angles and GRFs by not over-fitting, as shown in Figure 2.

ESNs were used in the literature to calculate joint angles and GRFs in patients with osteoarthritis stepping down [1]. [7] was used as a guide to understand ESN framework, reservoir size, spectral radius, and the remaining hyper-parameter selection. ESNs require a reservoir computing framework and additional functions to work. Pytorch-esn was used to implement the ESN which interfaced with the Trainer module in our code.

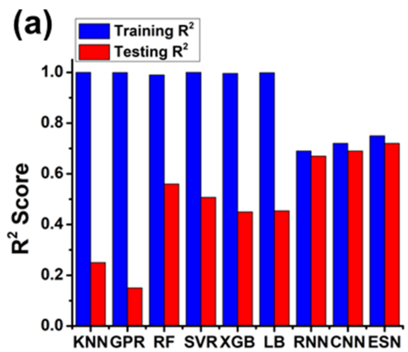


Figure 2: RNN, CNN, and ESN outperforming traditional ML approaches on time series data in [1]

3 Methods

3.1 Data Collection

We trained preliminary models on a validated dataset of EMG and GRF measurements which included video, marker data, and inverse dynamics as well [10]. While this gave us reason to believe that we could model GRFs as a function of EMG, the dataset had limitations that prohibited us from making generalizable and robust models. The set contained data on ten subjects doing seven activities for about five to ten seconds each and only half of that window contained activity. The diversity in subjects movement and our inability to generalize between motions meant that we did not have enough data to train robust models for each motion.

We elected to capture data on ourselves using resources in the Human Performance Lab at Stanford University. Three male subjects in their 20s walked for 20 minutes each at speeds of 1 to 1.2 m/s. We placed 14 EMG sensors on the lower limbs. We placed sensors over the right and left soleus, gastrocnemius, tibialis anterior, biceps femoris long head, vastus medialis, vastus lateralis, and the right semitendinosus and rectus femoris. EMG signal was measured with Delsys Corp wearable sensors and GRFs were measured with a force-instrumented treadmill. The setup can visualized in figure 3. Each subject has a typical gait pattern although one subject reported soreness and fatigue prior to the capture session. EMG was filtered with a 12Hz Butterworth filter and normalized to the maximum activation exhibited in a particular subject during walking. EMG was captured at 50Hz and GRFs were captured at 2kHz.

3.2 Model Training

We trained each model architecture on this data set. For each architecture, we produced four models that varied in the values they were trying to predict and which data was held out. Two were trained to predict GRFs on the right foot and two were trained to predict the left. In each of those

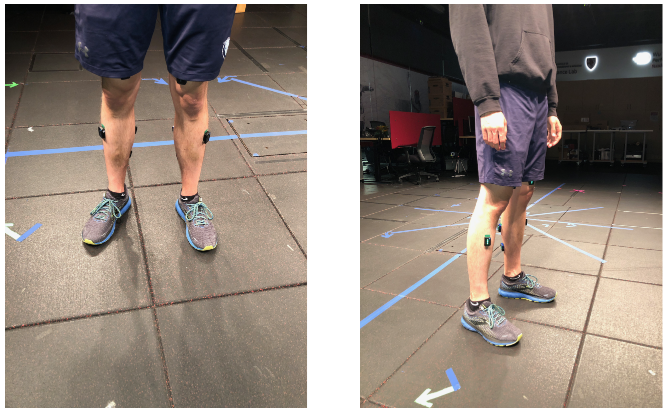


Figure 3: Collecting EMG Data in Human Performance Lab

groups, one was trained on 60% of each subject's data with 40% held out for validation and test. The other was trained on two subjects' data in entirety with the third subject being held out for validation and test. At training time we ran our model with a batch size of 32 and calculated loss with Mean Squared Error (MSE). We updated our model with PyTorch's Adam optimizer.

During training, we noticed that models would frequently fall into a local min of predicting 0 and we can see that each of the right and left curves is, indeed, 0 for about 40% of the gait cycle. To account for this, we ignored batches where every element was 0 with 99% probability.

4 Results and Discussion

In an effort to validate code, prior to training on the entire dataset, we looked at a subset of data. From a quick visual evaluation, models were incredibly accurate in their estimation of GRFs. Figure 4 shows a very strong fit. Admittedly this is a plot of model performance on the data it was trained on. Unfortunately, these results were not replicated when trained on the entire dataset and evaluated on a validation set. Because ESNs were not implementable under the same framework used for the other three architectures, their results are presented separately.

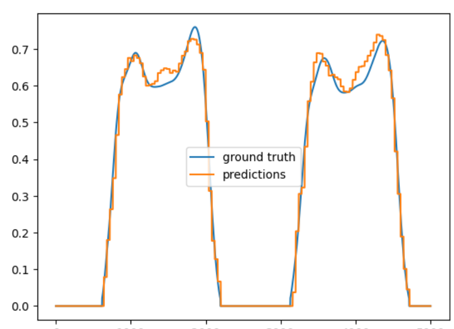


Figure 4: Models learn well when trained on small amounts of data

4.1 Learning Curves

Most models, excepting the fully connected MLP reduced cost over the course of their training. While this was encouraging, the performance was not reflected in observably satisfying accuracy. The model that didn't learn fell into the local minimum of predicting zero every time and it's easy to believe that it could be retrained to exhibit similar learning to other models.

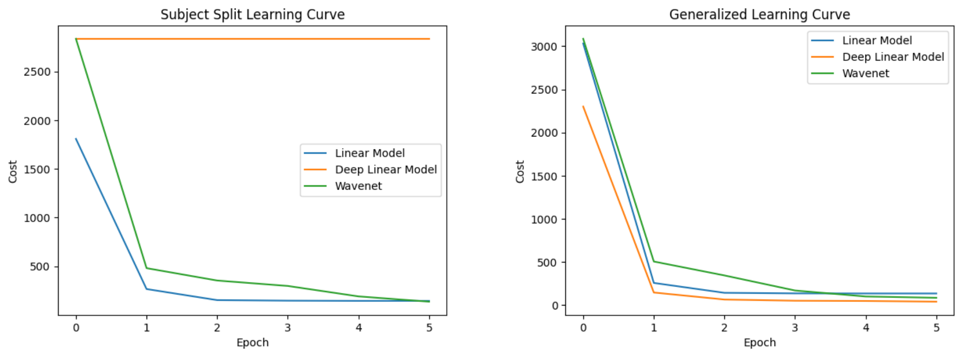


Figure 5: Learning Curve from Even Split and Generalizer Holdout

4.2 Time performance

Model GRF predictions resembled the general shape of GRFs; however, a significant differentiating factor between model architectures was the time to produce a series of GRF predictions. In practice, a GRF predictive model would be applied in real time; therefore, a faster model is preferred. A linear architecture presented the fastest run time compared to the neural network architectures as depicted in Table 1.

Table 1: Time performance

Model Architecture	Time (s) for forward run (10k samples)
Linear	4.795
Deep Linear	7.328
CNN	11.187

4.3 Simple Linear and Fully Connected Performance

The simple linear model in Figure 6 is the fastest model but presents significant noise compared to the simple fully connected neural network in Figure 7. The shallow neural network captures the small nonlinearities in the relationship between EMG and GRF data streams. In practice, software could implement filtering when the GRF prediction is below a threshold to eliminate the additional noise at low GRF predictions.

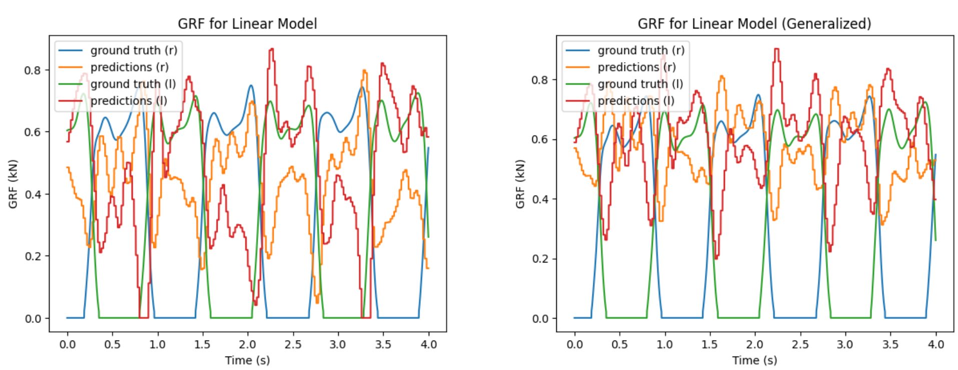


Figure 6: Performance of Single Layer Linear Model

4.4 WaveNet Performance

The convoluted architecture of the WaveNet model is depicted in Figure 8. The plots demonstrate the ability for the WaveNet architecture to generate reasonable GRF predictions; however, the model did

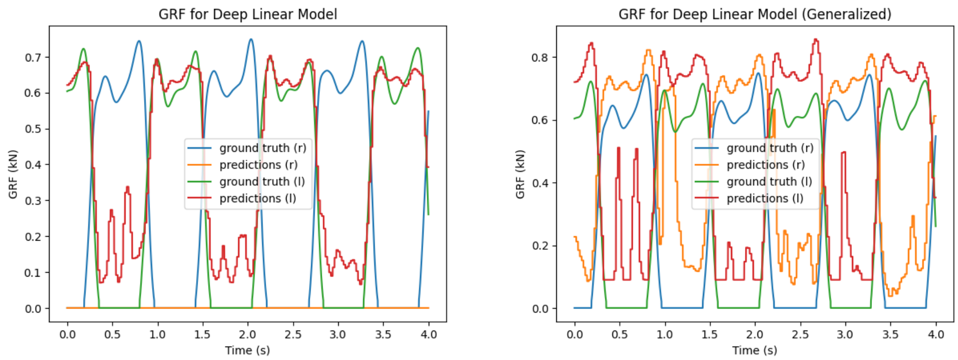


Figure 7: Performance of Deep Linear Model

not predict the GRFs as well as the shallow neural network in Figure 7. The complex convoluted neural network attempted to learn the GRF pattern with stacked convolutional layers with small kernels. This approach does not emphasize a direct relationship between EMG and GRF compared to the linear layers in the previously mentioned approaches.

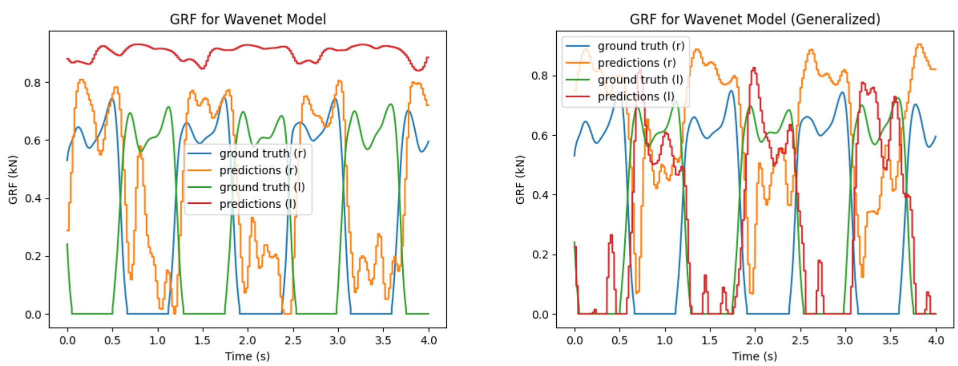
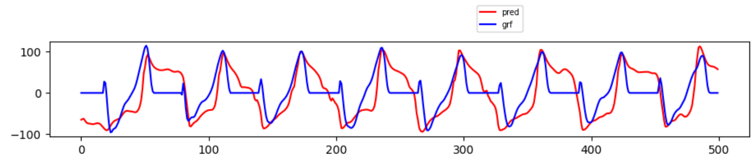


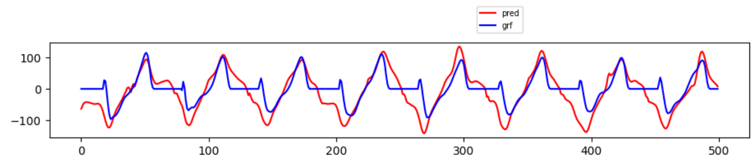
Figure 8: Performance of CNN

4.5 ESN Performance

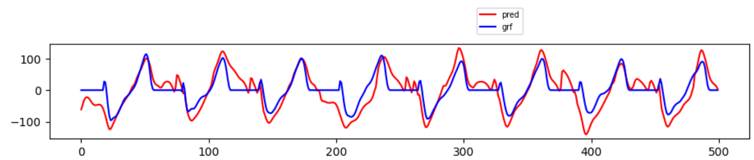
For peak fitting, The best result of the PyESN-based implementation was an RMSE of 32, using either 2000 or 5 reservoir neurons and a spectral radius of 1.5 (shown in Figure 9a). For fitting flat regions, the best result was from, a reservoir of 100 and a spectral radius of 90. However, fitting peaks and fitting zeroes have an inverse correlation by inspection. Even the best models calculated by RMSE oscillated jaggedly around the flat regions - minimizing error but not actually fitting very well aside from peaks. Despite the conventional wisdom that larger reservoirs lead to better performance, the lowest RMSE values were consistently at lower reservoirs (5-50 total nodes). These small models train in under 30 seconds - whereas 4000 node reservoirs took hours (and had worse performance, as shown in table 2 below). One possible explanation for this is that simple time-series data only needs a small reservoir, as it also serves as a short-term memory for the data - which is simple.



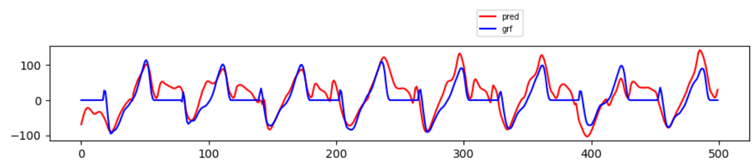
(a)



(b)



(c)



(d)

Figure 9: GRF Predictions for the ESN Model, corresponding to samples 1-4 in the table below.

Given that Cooper was unable to correctly implement an ESN that matched the PyTorch implementation, it isn't possible to evaluate the performance relative to the other models. Regardless, the ESN was a viable option - provided that hyper-parameters were tuned sufficiently. Despite not being easy to fit into our PyTorch framework, we expect that this model is still a viable option for this task as a result of the work done by [1], shown in Figure 2. However, the complexity is only merited when training to predict multiple outputs (Joint angles, GRFs, etc).

5 Future Work

This report demonstrates an accurate model in predicting real-time GRFs for walking. Current force place GRF measuring devices provide reliable GRFs for any task. Therefore, future work should explore developing models for GRF prediction from EMG data across a variety of movement-related tasks. These tasks may include the following: running, jumping, asymmetric walking, squatting, and gait patterns associated with diagnosed musculoskeletal diseases. Subsequently, approaches should be investigated to combine such models to yield a single platform that may provide real-time GRFs without the necessary force plate setup. One such combination tactic may be to first predict the movement pattern and then apply the appropriate model to a single EMG window accordingly.

In addition, refining the ESN and fully integrating it with the PyTorch framework would allow for quicker iteration, as well as more nimble training splits between different subjects. This would enable the ESN to be evaluated for its ability to generalize across subjects, as well as encompass both left and right foot data at the same time. The main bottleneck to this is creating a Training class that was able to pass parameters such as washout (which operated similarly to WaveNet looking back at n samples) all the up through the model.

Table 2: ESN RMSE Performance with Hyperparameter Exploration

Reservoir Neurons	Spectral Radius	Random Seed	RMSE
250	1.5	42	57.8583
500	1.5	42	59.5360
1000	1.5	42	49.5854
2000	**1.5**	**42**	**32.2617**
4000	1.5	42	41.2431
500	3	42	54.7051
500	1	42	47.5439
500	0.5	42	46.4272
40	1.5	42	55.5682
20	1.5	42	51.7118
5	**1.5**	**42**	**32.7038**
100	**90**	**42**	**26.6790**

6 Contributions

The report was written with equal contributions from the team.

Ben contributed to the initial exploration with the WaveNet and attempts to generate a simple fully connected MLP for the GRF prediction. Ben facilitated EMG and GRF data collection in the Stanford HPL with teammates as well as post processing the data for appropriate model development.

Quincy contributed to reading in the processed EMG and GRF data. Quincy developed the most successful linear regression, fully-connected, MLP, and convolutional architectures to predict GRF forces from EMG inputs. Quincy established the framework for plotting the GRF outputs with the ground truth labels as well.

Cooper contributed to the development of the echo state network GRF prediction as well as a literature review and exploration/analysis of previous methods.

References

[1] Richard Byfield, Matthew Guess, Kianoosh Sattari, Yunchao Xie, Trent Guess, and Jian Lin. Machine learning full 3-d lower-body kinematics and kinetics on patients with osteoarthritis from electromyography. 5:100088.

[2] Reed D Gurchiek, Ryan S McGinnis, Alan R Needle, Jeffrey M McBride, and Herman van Werkhoven. The use of a single inertial sensor to estimate 3-dimensional ground reaction force during accelerative running tasks. *Journal of Biomechanics*, 61:263–268, 2017.

[3] Ernest Nlandu Kamavuako, Dario Farina, Ken Yoshida, and Winnie Jensen. Estimation of Grasping Force from Features of Intramuscular EMG Signals with Mirrored Bilateral Training. *Annals of Biomedical Engineering*, 40(3):648–656, March 2012.

[4] Zoe Mass Kokolevich, Erik Biros, Oren Tirosh, and Jacqueline Elise Reznik. Distinct Ground Reaction Forces in Gait between the Paretic and Non-Paretic Leg of Stroke Patients: A Paradigm for Innovative Physiotherapy Intervention., nov 2021.

[5] Jie Liu, Xiaoyan Li, Guanglin Li, and Ping Zhou. EMG Feature Assessment for Myoelectric Pattern Recognition and Channel Selection: A Study with Incomplete Spinal Cord Injury. *Medical engineering & physics*, 36(7):975–980, July 2014.

[6] David G Lloyd and Thor F Besier. An EMG-driven musculoskeletal model to estimate muscle forces and knee joint moments in vivo. *Journal of Biomechanics*, 36(6):765–776, June 2003.

[7] Mantas Lukoševičius. A Practical Guide to Applying Echo State Networks. In Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller, editors, *Neural Networks: Tricks of the Trade: Second Edition*, Lecture Notes in Computer Science, pages 659–686. Springer, Berlin, Heidelberg, 2012.

[8] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. WaveNet: A generative model for raw audio.

[9] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio, 2016. cite arxiv:1609.03499.

[10] Scott Uhlrich, Antoine Falisse, and Jennifer Hicks. Dataset: Lab validation with videos.