1J2-1

# Rapid response amplitude control of high-power ultrasonic transducer using deep reinforcement learning

Tatsuki Sasamura[1†], Yanbo Wang[2] and Takeshi Morita[1,2*] ([1]Grad. School of Frontier Sciences, The Univ. of Tokyo; [2]Grad. School of Eng., The Univ. of Tokyo)

## 1. Introduction

High-power resonant-type ultrasonic transducers play a vital role in various industrial applications. They are essential for processes including ultrasonic welding, jointing, and cutting, as well as ultrasonic motors. These applications require rapid control due to the short processing time.[1)] The driving frequency is commonly used to control vibration; however, since the frequency has a nonlinear relationship with vibration, conventional linear control, such as PID control, often results in suboptimal control, which sacrifices quick response. In addition, the resonance frequency can shift under high-power conditions, which causes the jump phenomenon and hysteresis to the frequency characteristics, making control even more difficult.

To overcome these complexities, this work proposes a deep reinforcement learning system that effectively controls high-power transducers.

## 2. Control system

The bolted Langevin transducer (BLT) (FBL28302SSF-FC, Fuji Ceramics Corporation) was employed as the control subject. This transducer features a 30 mm diameter and 91 mm length.

The system shown in **Fig. 1** is controlled by a microcontroller (ESP32-S3-DevKitC, Espressif Systems) on a control board, which commands direct digital synthesizers (DDSs) for sinusoidal output, which is used for the driving signal. The DDS driving signal is amplified by bipolar amplifiers (HSA4052, NF Corp.) to power the BLT at 60Vpp.

The driving current is converted to a voltage signal via a transformer, which is then transformed into two direct voltage signals by a phase-sensitive detector (PSD). The PSD produces output voltages proportional to active and reactive driving currents.

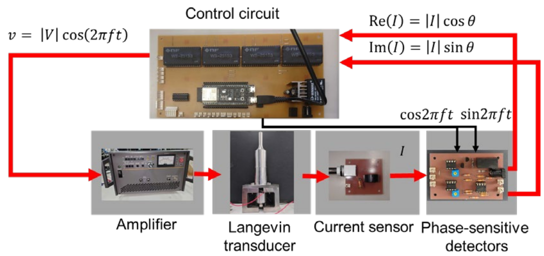


Fig. 1 Experimental setup.

## 3. Characteristic measurement of BLT

### 3.1 Quasi-static measurement

**Figure 2a** shows the admittance of BLT under 4 Vpp voltage, measured by an impedance analyzer (4249A, Keysight Technologies). It indicates resonance frequency at $f_r$ =28.669 kHz, half-width at 3 Hz. thus, the quality factor at $Q$ =9600. The theoretical time constant of the vibration growth can be calculated as $\tau = Q/\pi f_r$ =107 ms.

Meanwhile, the admittance under 60 Vpp voltage was measured using the system described in the previous section, as shown in **Fig. 2b**, which shows the jumping phenomenon and hysteresis.

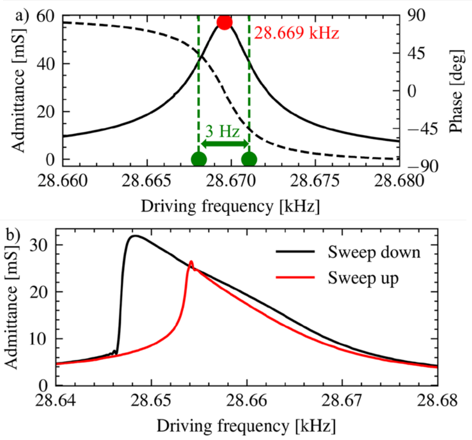


Fig. 2 Admittance of BLT a) 4Vpp, b) 60 Vpp

### 3.2 Integral control response

To show the difficulty of the frequency control, the driving current was controlled by integral control as formulated with

$$f = f_0 + K_i \int (I - I_{target})dt$$

where $K_i$ =700 Hz/A・s, $f_0$ =28.7 kHz, and $I_{target}$=0.3,0.5,0.7,0.9 A, as shown in **Fig 3**.

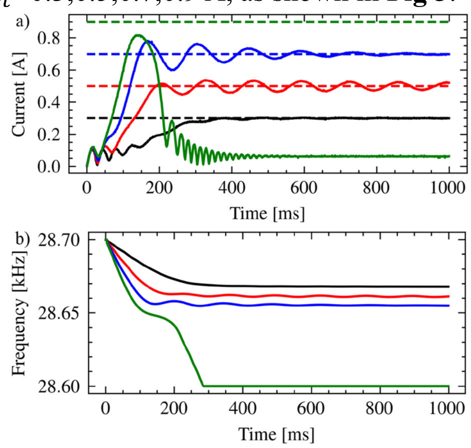


Fig. 3 Response of the integral control: a) current amplitude(bold) and target (dashed), b) frequency.

Although the integral control converged in 400 ms without overshooting at 0.3 A target, it suffered continuous ringing at 0.5 A and completely diverged at 0.9 A. These results demonstrate the difficulty of the control due to the transducer's nonlinearity.

## 4. Deep reinforcement learning control

### 4.1 Formulation of Malkov decision process

Reinforcement Learning is a type of machine learning that optimizes a controller's behavior within a Markov Decision Process (MDP). The current control is modeled as an MDP, illustrated in **Fig. 4**.

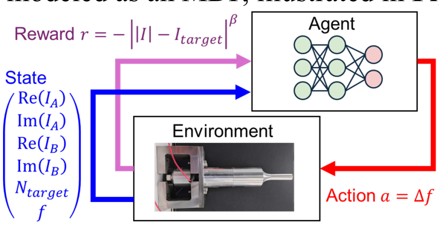


Fig. 4 Malkov decision process formulation

In an MDP, the controller (agent) interacts with the environment over several steps. At each step $n$, the agent receives a state $s_n$ and takes an action $a_n$ based on that state. The environment then provides a reward $r_n$ and updates the state to $s_{n+1}$. This process repeats every 1 ms, creating a sequence of states, actions, and rewards known as an episode.

The ultimate goal of reinforcement learning is to maximize the cumulative reward in the episode, defined as

$$R = \sum_n \gamma^n r_n$$

where $\gamma$ is the discount factor in the range of $0 < \gamma < 1$. In this work, the reward is defined as

$$r = -\left|I_0 - I_{target}\right|^{\beta}$$

where $I_{target}$ is the target, $\beta = 0.5$ is constant.

Meanwhile, the state has four components: the active and reactive part of the driving current $\mathrm{Re}(I)$ and $\mathrm{Im}(I)$, the target amplitude $I_{target}$, and the driving frequency $f$. Then, the agent decides the action $a$, which determines the increment of the driving frequency as $\Delta f = Da$, where $D$ =10 Hz is the frequency step range.

### 4.2 Soft actor-critic networks [3)]

The agent's action in this system is determined by a probability distribution $\pi(a|s)$, implemented through an actor neural network. The input layer receives the state, and through a hidden layer, the output layer produces two outputs: mean $\mu$ and the logarithm of variance $\ln \sigma$. A random number $u$ is then sampled from a Gaussian distribution $N(\mu, \sigma)$, and the action is calculated as $a = \tanh u$.

Training is conducted over 1000 episodes, each with 500 steps, with a randomly set target current between 0.1 A and 1 A.

## 5. Control results

After training, the agent controls the frequency using the trained parameters for target currents $I_{target}$=0.3, 0.5, 0.7, 0.9 A. The driving current response is shown in **Fig. 5**.

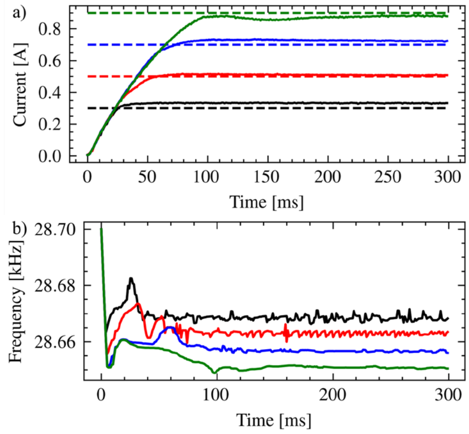


Fig. 5 Result of the DRL control: a) current amplitude(bold) and target (dashed), b) frequency.

The responses almost follow the first-order system's response, reaching the target in 100 ms in the case of 0.9 A target current. This is almost equal to the theoretical time constant, which indicates that the DRL frequency control optimally tunes the frequency against the nonlinear vibration system.

## 6. Conclusion

This work demonstrates the effectiveness of a deep reinforcement learning (DRL) approach for controlling high-power ultrasonic transducers using only frequency modulation. By formulating the control problem as a Markov Decision Process (MDP), the DRL system successfully adapts to the nonlinear characteristics inherent to high-power ultrasonic vibrations. The results show that the DRL control achieves near-optimal performance, with response times closely matching the theoretical limits. This approach offers a robust solution for industrial applications requiring rapid ultrasonic transducer control without increasing the complexity of the driving circuit. Our future works include 1) validating the DRL method under external conditions such as temperature and boundary conditions and 2) applying the DRL method to different transducers, including bending or sheer transducers and ultrasonic motors.


## Acknowledgment

This work was supported by the JSPS DC2 program for T. Sasamura under Grant 23KJ0571.


## References


1) J. Wang, J Jiang, F. Duan, F. Zhang, W. Liu, X. Qu, IEEE Trans. Ind. Electron. 67 (2020) 6864–6873.
2) S. Mojrzisch, J. Wallaschek, J. Intell. Mater. Syst. Struct. 24 (2013) 745–752.
3) T. Haarnoja, A. Zhou, P. Abbeel and S. Levine, arXiv [cs.LG], 2018.