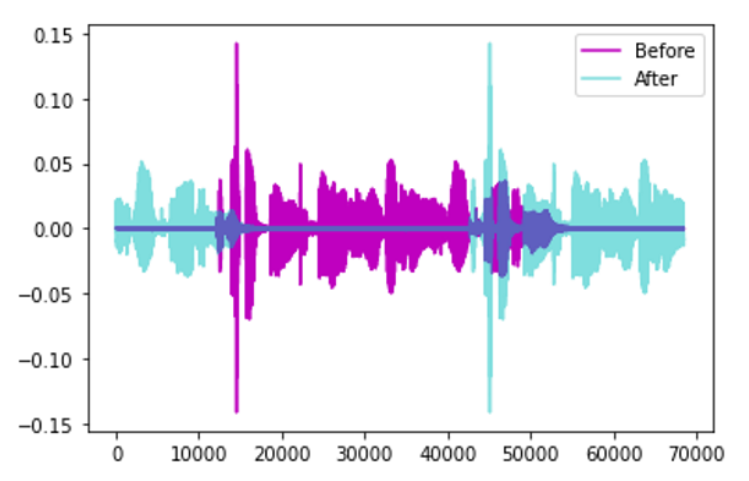
**Tiền xử lý dữ liệu**

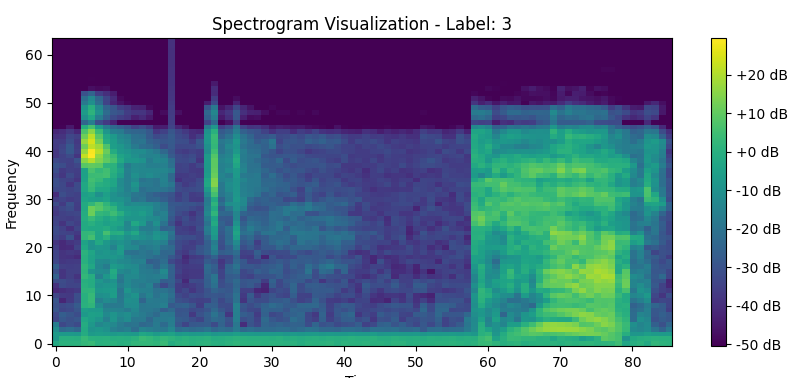
Trước khi đưa dữ liệu vào model ta cần thực hiện một số bước tiền xử lý, quá trình này được thực hiện một cách tự động cùng lúc với việc đọc files audio

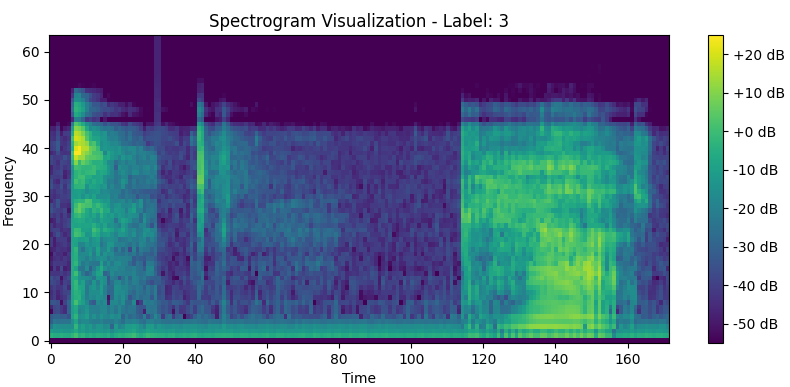
Các bước tiền xử lý lần lượt như sau:

* **Đọc file audio**
* **Chuyển đổi về cùng kênh:** một vài audio có thể ở dạng một kênh (mono), một số khác ở dạng hai kênh (stereo). Vì model chỉ chấp nhận dữ liệu cùng kích thước nên ta chuyển hết về dạng mono.
* **Chuyển về cùng Sample rate:** Vì các file audio có thể có các tần số lấy mẫu khác nhau mà model chỉ chấp nhận dữ liệu cùng kích thước nên ta chuyển hết về 16000Hz.
* **Thay đổi chiều dài các file audio về chung một giá trị:** file có chiều dài lớn hơn giá trị quy ước sẽ được cắt bớt đi, file có chiều dài nhỏ hơn sẽ được thêm khoảng lặng.
* **Thực hiên kỹ thuật Augmentation là Time-Shift:** kỹ thuật này giúp mô hình học được các biến thể của âm thanh khi chúng bị dịch chuyển theo thời gian, làm cho mô hình ít bị phụ thuộc vào vị trí chính xác của tín hiệu trong đoạn âm thanh và có thể giúp cải thiện độ chính xác khi làm việc với các tín hiệu có thể bị dịch chuyển trong thực tế.



* **Chuyển đổi sang Mel Spectrogram:** chuyển tín hiệu âm thanh thành Mel-spectrogram và sau đó chuyển đổi các giá trị biên độ của nó thành đơn vị decibel (dB). Các bước này giúp mô hình học máy dễ dàng nhận diện đặc trưng tần số của âm thanh, đặc biệt là trong các bài toán nhận diện âm thanh, giọng nói hay phân loại tiếng ồn. Mel-spectrogram thường được sử dụng trong các mô hình học sâu vì nó phản ánh các đặc trưng quan trọng của âm thanh, đặc biệt là trong môi trường âm thanh không gian nhỏ.





* **Thực hiện kỹ thuật Augmentation là Time and Frequency Masking đối với Mel Spectrogram:** tăng cường dữ liệu cho phổ Mel bằng cách sử dụng Frequency Masking và Time Masking để che khuất một phần của phổ tần số và thời gian. Điều này giúp mô hình học máy học cách nhận diện âm thanh ngay cả khi có sự thiếu hụt hoặc biến dạng trong dữ liệu, từ đó tăng khả năng tổng quát của mô hình khi gặp phải những tín hiệu âm thanh không hoàn chỉnh hoặc có nhiễu.

