
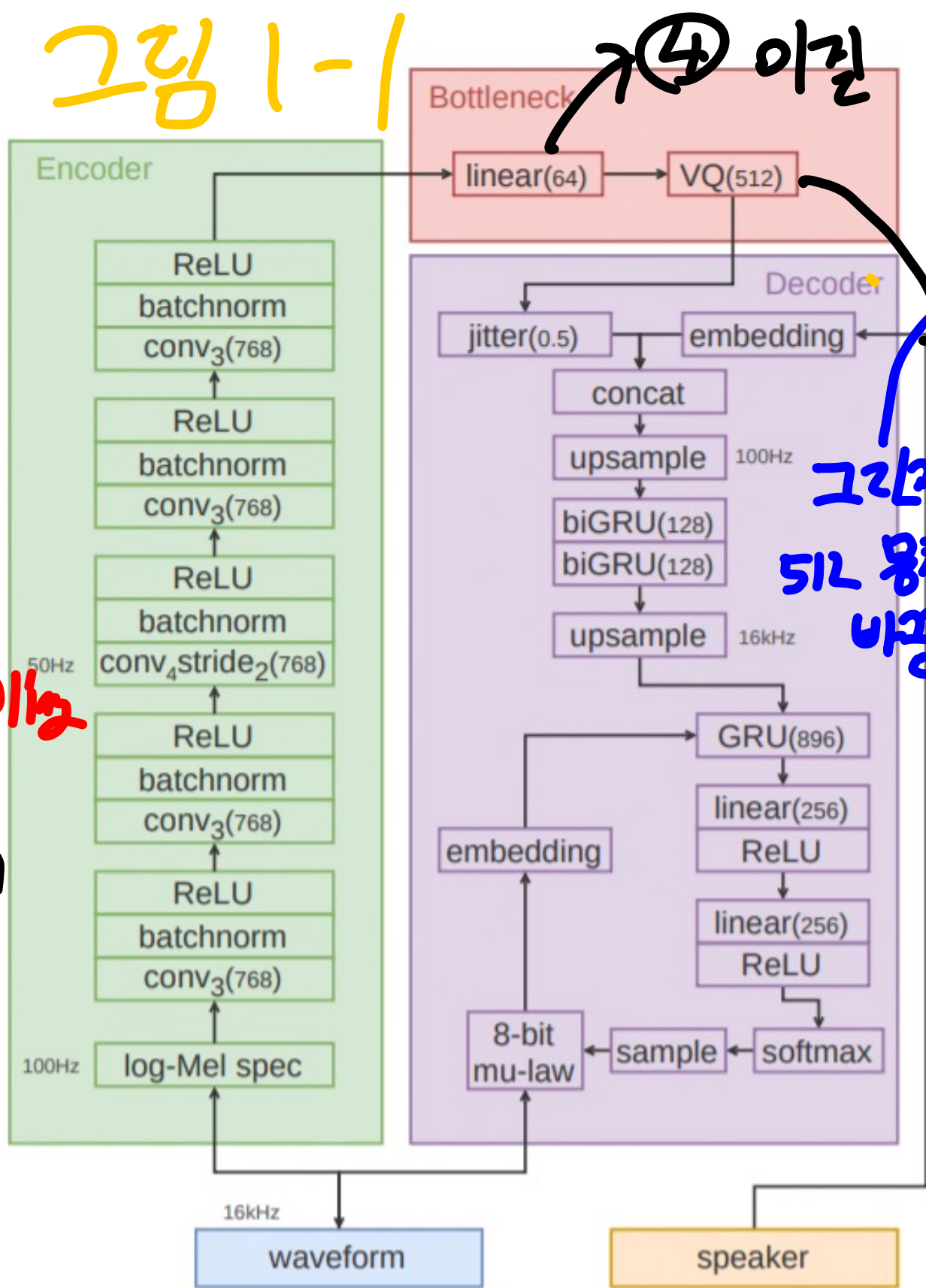


그림 1-1

- ③ 계층적
0.3
2'  벡터로
포함
- ② 574 ||
convolution
layer 개수
- ① log mel
변이면



- ④ 이걸 continuous
하게 변환
(시간순)
- ⑤ VQ에서
벡터 양과
바꾸는
이유
- 그런
512 용도로
바꾸는
- C_t 벡터로
|| 포함
code book
시계열 처리,
학습?
 C_{t+1} 예측
= 4
시간순 배열.

그림 2

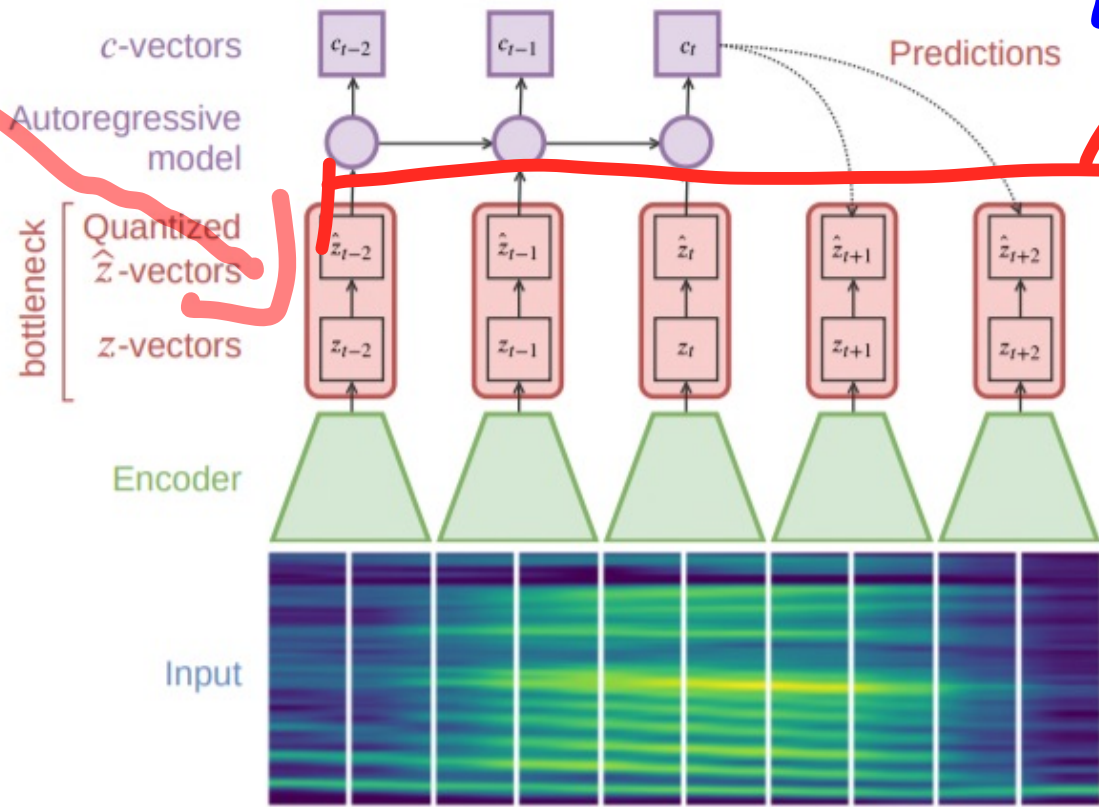
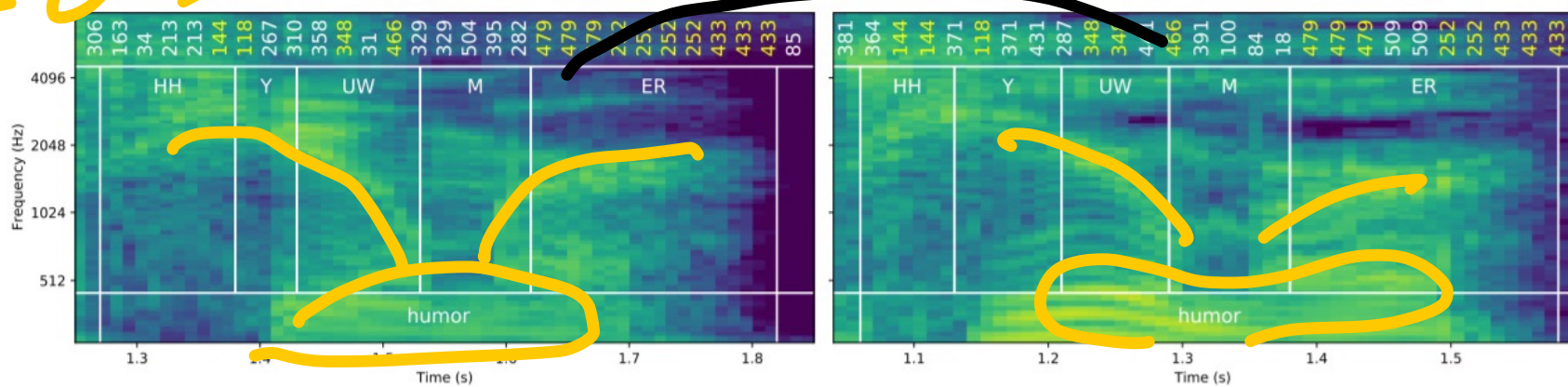
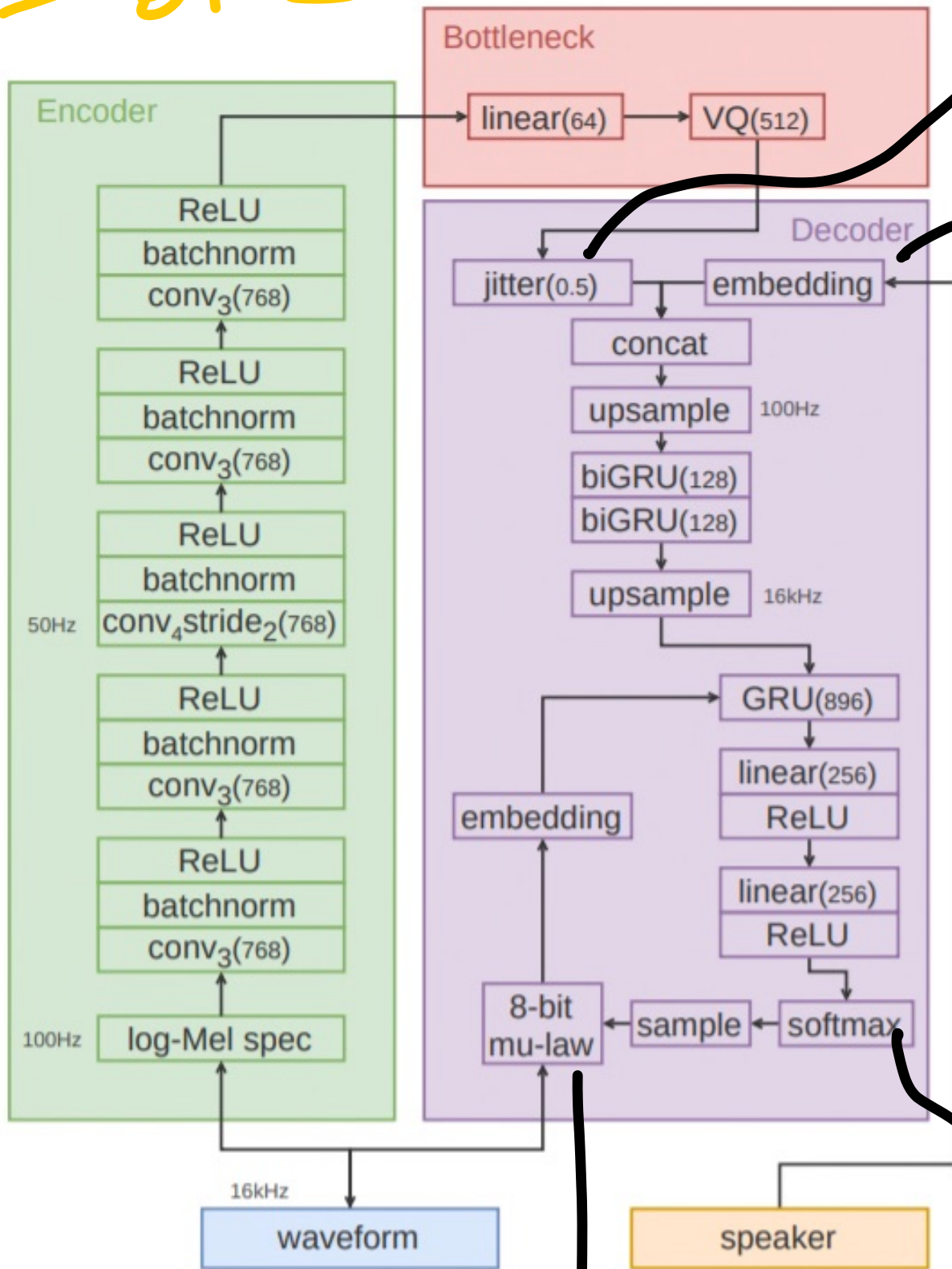


그림 3



이 부분이
공통
부분

그림 1-2



⑤ 과적합 방지

⑥ 사용자 목소리 특징 뽑아서 벡터화

목소리 합성과 관련된 여러 가지 문제점을
ㄱ ㄱ

multi class 분류

음성 8비트,
16비트 맞춰주는 작업.