대화 맥락을 반영한 백채널 예측 모델

최용석°[1], 박요한[1], Wencke Liermann[2], 이공주[1]
충남대학교
[1]{yseokchoi,happy115012,kjoolee}@cnu.ac.kr
[2]wliermann@o.cnu.ac.kr

Conversation Context-Aware Backchannel Prediction Model

Yong-Seok Choi°, Yo-Han Park, Wencke Liermann, Kong Joo Lee
Chungnam National University

요 약

백채널은 화자의 말에 언어 및 비언어적으로 반응하는 것으로 상대의 대화 참여를 유도하는 역할을 한다. 백채널은 보편형 대화 참여와 반응형 대화 참여로 나뉠 수 있다. 보편형 대화 참여는 화자에게 대화를 장려하도록 하는 단순한 반응이다. 반면에 반응형 대화 참여는 화자의 발화 의도를 파악하고 그에 맞게 반응하는 것이다. 이때 발화의 의미를 파악하기 위해서는 표면적인 의미뿐만 아니라 대화의 맥락을 이해해야 한다. 본 논문에서는 대화 맥락을 반영한 백채널 예측 모델을 제안하고 예측 성능을 개선하고자 한다. 대화 맥락을 요약하기 위한 방법으로 전체 요약과 선택 요약을 제안한다. 한국어 상담 데이터를 대상으로 실험한 결과는 현재 발화만 사용했을 때보다 제안한 방식으로 대화 맥락을 반영했을 때 성능이 향상되었다.

주제어: 백채널 예측, 발화 단위, 대화 맥락

1. 서론

사람들 간의 대화에서는 언어 및 비언어적 상호작용을 통해 자연스러운 대화가 이루어진다[1]. 이러한 상호 작용 중 하나인 백채널은 화자의 발화를 방해하지 않으면서 대화 주제에 대한 관심과 의견을 표현하는 것이다[2]. 대화 도중에 적절하게 백채널을 사용하는 것은 화자의 더욱 풍부하게 대화를 유도할 뿐만 아니라 신뢰를 구축하는 데도 도움이 된다[3]. 특히 상담 대화에서 상담사의 백채널은 내담자의 말을 유도하고 공감을 표현하여 신뢰 관계를 형성하는데 중요한 역할을 한다[4].

백채널은 크게 두 가지로 나뉜다[5]. 첫 번째는 보편형 대화 참여로 화자가 발화한 내용을 인정하거나 이해하여 화자가 대화를 계속 이어갈 수 있도록 장려하는 것이다. 두 번째는 반응형 대화 참여로 청자가 화자의 발화를 기반으로 응답하는 것이다. 보편형 대화 참여는 대화 맥락에 관계없이 발생될 수 있지만 반응형 대화 참여는 화자가 발화한 것을 기반으로 응답을 생성해내기 때문에 발화의 의미를 파악하는 것이 중요하다[6]. 대화 도중에 생성된 발화는 현재까지의 흐름과 감정 등을 종합적으로 고려한다[7]. 그러므로 발화의 의미를 파악하기 위해서는 표면적인 의미뿐만 아니라 대화의 맥락을 이해하는 것이 필요하다.

이전 백채널 예측 연구들은 음성 신호와 텍스트를 모델의 입력으로 사용하였다. 이때, 텍스트 입력은 고정된 길이의 단어를 입력하거나[8, 9] Inter Pausal Unit (IPU) 단위로 입력하였다[10]. 이것은 현재의 발화만 고려하고 있다. 그러나 백채널 예측은 이전 발화의 내용을 이해하고 대화의 맥락을 고려하는 것도 필요하다.

본 논문에서는 백채널 예측 모델에 대화 맥락을 반영하기 위한 두 가지 방법을 제안한다. 첫 번째는 대화의 흐름을 순차적으로 요약하는 것이다. 두 번째는 정보가 생략된 현재 발화의 정확한 의미를 파악하기 위해 이전 발화에서 선택적으로 내용을 요약하는 것이다. 이 두 가지 방법을 통해 대화 맥락을 반영한 백채널 예측 모델을 제안한다.

2. 관련연구

백채널 예측 모델은 주로 텍스트나 음성 신호를 사용한다. 텍스트를 정보를 활용할 때에는 자동 음성 인식기의 성능이나 출력 시간 지연 문제의 제한 사항이 있다[8, 10]. [11]은 음성 신호가 텍스트에 비해 백채널 예측에서 더 큰 역할을 하였다. 이러한 이유로 초기 모델은 음성 신호만 사용하였다[12].

그 후에 텍스트를 함께 활용하는 모델이 제안되었다. [8]는 텍스트와 음성 신호를 함께 활용하였을 때 모델의 성능이 향상됨을 보였다. 텍스트 정보를 추출하는 방법은 단어 임베딩[8, 11]이나 사전 학습된 언어 모델[9]이 주로 활용된다. 이때 텍스트의 단위 및 길이에 대한 다양한 연구가 이루어졌다.

텍스트 입력 단위의 다양한 접근 방식 중에서 [11]은 음성 신호와 텍스트를 함께 사용하면서 동일한 시간에 발생한 단어를 추출하여 입력으로 사용하였다. [10]은 200ms의 묵음을 기준으로 발화를 분리하고 이것을 Inter Pausal Unit (IPU)으로 정의하였다. 또 다른 접근 방식은 고정된 길이의 단어를 사용하는 것이다. [8]은 고정된 길이의 단어를 추출하고 CNN을 사용하였다. 이때 입력 길이가 짧을 때 성능이 더 좋음을 보였

다. 반면에 [9]은 고정된 길이의 단어를 사전 학습 모델 BERT에 입력으로 넣어주었다. 실험을 통해 입력 단어가 많을수록 성능이 향상됨을 보였다.

백채널 중에 반응형 대화 참여는 대화 맥락에 대한 고려가 필요하다. 하지만 이전 연구들은 백채널 예측에 있어 대화 맥락을 고려하는 방법은 제안되지 않았다. 본 연구에서는 백채널 예측을 목적으로 현재 발화에 음성 신호와 텍스트를 고려하는 것 뿐만 아니라 다중 턴 대화에서 이전 발화들을 활용해 맥락을 고려해보고자 한다.

3. 백채널 예측 모델

3.1 시나리오

본 논문은 다중 턴 대화에서 대화의 맥락을 고려한 백채널 예측을 목표로 한다. 백채널 예측 시나리오는 두 명의 사람이 대화하는 것으로 가정한다. 대화는 T개의 발화로 이루어져 있으며($D = \{U_0, U_1, ..., U_T\}$), 각 발화는 N개의 단어로 구성된다($U_t = \{w_0, w_1, ..., w_N\}$). 백채널은 현재 발화 도중 또는 끝에 발생될 수 있다.

3.2 모델 구조

백채널 예측 모델의 구조는 그림 1과 같다. 이 모델은 [13]에서 제안된 모델 구조를 기반으로 한다. 모델은 음성 신호와 현재 발화, 이전 발화 정보를 활용하여 백채널을 예측한다.

3.2.1 음성 신호

음성 신호는 백채널 예측 모델에서 가장 기본적으로 사용된 정보이다[11]. 그렇기 때문에 음성 신호를 인코딩 하여 모델의 입력으로 사용하였다. 음성 신호 표현은 wav2vec 2.0[14]를 사용하여 고차원의 임베딩으로 인코딩하였다(E_A). wav2vec으로 생성된 임베딩은 음성 신호의 길이만큼 추출되기 때문에 평균값을 취하였다. 음성 신호는 이전 연구[8, 9]와 동일하게 백채널을 예측하는 시점으로부터 1.5초 전의 신호를 사용하였다.

3.2.2 현재 발화

백채널 예측에서 음성 신호는 중요한 역할을 하지만 텍스트 정보을 함께 사용하면 더 나은 성능을 보였다[8]. 따라서 현재 화자의 발화 텍스트를 백채널 예측에 활용한다. 현재 발화의 텍스트 표현은 BERT[15]를 사용한다. 이 과정에서 현재 화자를 식별하고 각 화자의 특성을 모델에 반영하기 위해 화자 임베딩($[Speaker]$)을 추가하였다. 현재 발화의 표현으로서 $[CLS]$ 토큰의 임베딩 사용한다(U_T).

3.2.3 이전 발화

다중 턴 대화에서는 두 가지 특징이 있다. 첫째, 화자의 발화를 정확히 이해하려면 이전 대화의 맥락을 이해해야 한다[16, 17]. 이는 맥락에 따라 발화의 의미가 달라질 수 있기 때문이다. 둘째, 다중 턴 대화에서는 종종 이전 발화에서 언급된 개념이나 개체를 생략하거나 대명사로 대체가 된다[18]. 그러므로 일부 정보가 생략된 현재 발화는 이전 발화로부터 누락된 정보를 추론해야 한다. 이 두 특징들은 화자의 발화를 이해하기 위해 이전 발화 등의 정보를 활용해야 한다는 공통점이 있다.

본 논문은 백채널 예측 모델에 이전 발화 정보를 활용한다. 먼저 최근 k개의 이전 발화를 임베딩하여 메모리에 저장한다. 이전 발화의 임베딩은 현재 발화와 동일한 방식으로 생성한다. 저장된 메모리의 임베딩을 활용하여 두 가지 방법으로 대화의 맥락을 요약하고자 한다.

첫 번째 방법은 전체 요약이다(C_{HOL}). 대화는 이전 발화와 관련해서 상호작용이 이루어진다. 따라서 대화의 전체 맥락을 이해하기 위해서는 각 발화를 순차적으로 요약해야 한다. 전체 요약의 모델은 GRU[19]를 채택하였다. GRU의 입력은 저장된 메모리의 이전 발화들의 표현들이다. 최종적으로 전체 요약의 표현은 GRU의 마지막으로 출력된 벡터를 사용한다.

두 번째 방법은 선택 요약이다(C_{SEL}). 정보가 부족하거나 생략된 경우에 현재 발화는 이전 발화로부터 정보를 보완하여야 한다. 하지만 이전 발화가 전부 관련된 것은 아니다. 이 관점으로부터 선택 요약은 [13]에서 제안한 TAA(Time-Aware Attention)을 적용한다(수식 1). 현재 발화는 이전 발화와 시간 차이가 벌어질수록 관련성이 떨어질 수 있다. TAA는 이러한 특성을 반영하여 기존 어텐션의 가중치에 시간 차이(D)의 역수를 곱하였다. 본 논문에서는 현재 발화를 쿼리(q)로서 사용하였고 k개의 이전 발화를 키(K)와 값(V)으로 사용한다. i는 현재 발화와 이전 발화의 시간 차이이다.

$$C_{SEL} = Softmax(\frac{qK^T}{D})V; D \in \mathbb{R}^k, D_i = i \quad (1)$$
$$q = W_Q U_T,$$
$$K = W_K U_{[T-k:T-1]},$$
$$V = W_V U_{[T-k:T-1]}$$

3.2.4 백채널 분류기

백채널 예측은 앞에서 언급한 네가지 정보를 활용한다. 구체적으로 음성 신호 임베딩(E_A), 현재 발화 임베딩(U_T), 전체 요약(C_{HOL}), 선택 요약(C_{SEL})을 모두 연결한 뒤 백채널 분류기의 입력으로 사용한다. 분류기는 여러 개의 선형 계층으로 구성하였다.

4. 실험 및 평가

4.1 데이터

실험은 ETRI에서 구축한 한국어 심리 상담 데이터를 사용하였다. 이 데이터는 약 32시간에 걸친 40개의 심리 상담 대화로

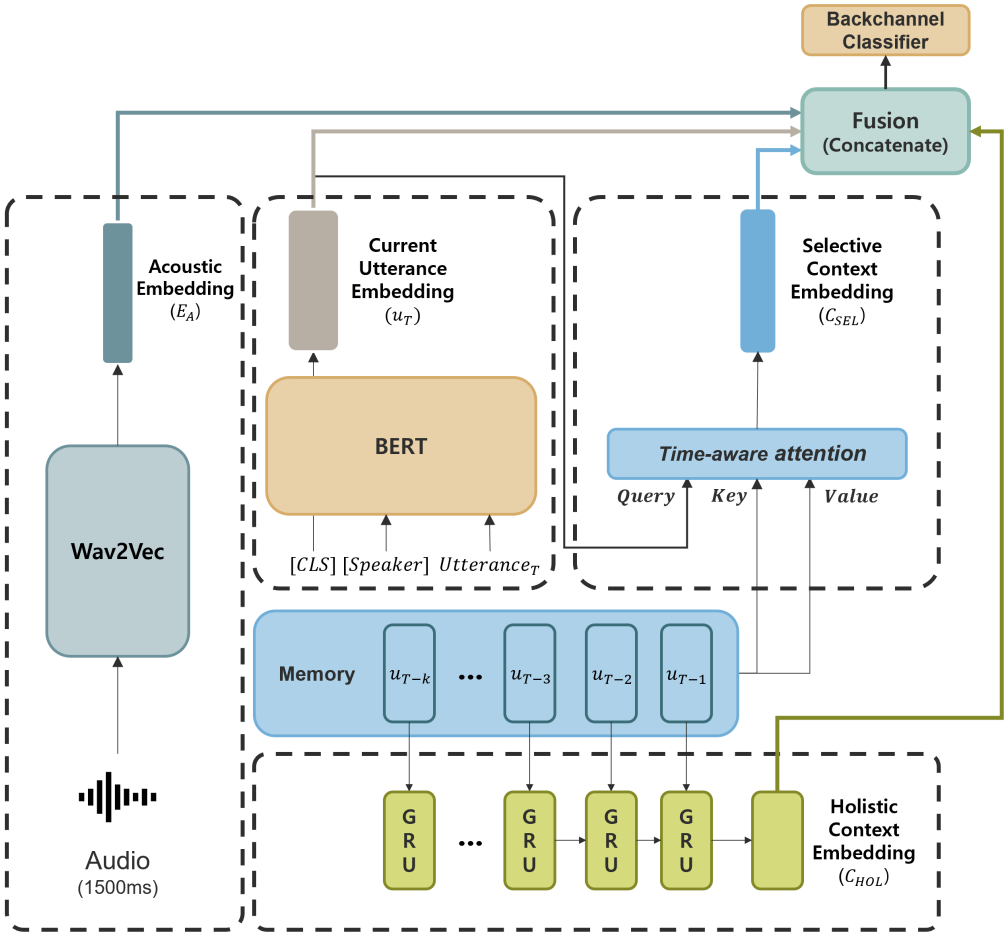

그림 1. 백채널 예측 모델 구조

구성되어 있다. 각 대화는 상담자와 내담자로 2명이 참여한다. 이 데이터는 오디오 및 전사문을 포함하고 백채널에 대한 라벨링도 함께 부착되어 있다.

백채널은 보편형 대화 참여인 Continuer와 반응형 대화 참여인 Understanding, Empathetic으로 분류되며 각각은 다음과 같이 정의된다. Continuer는 대화에 집중하고 있다는 표현에 대한 응답이다. 주로 '응', '네' 등과 같은 간단한 반응을 포함한다. Understanding은 대화에 집중하는 동시에 화자의 발화를 이해했음에 대한 반응이다. 주로 발화가 끝난 이후에 나타나는 경향이 있다. Empathetic은 발화 내용 뿐만 아니라 화자의 감정과 의도를 이해하고 이를 적극적으로 반영하여 응답하는 것이다. '오 ', '아이고 저런' 등과 같은 반응으로 화자의 발화 내용에 대한 감탄, 동의, 놀람의 의미를 가지는 반응이 포함된다.

이전 연구들[8, 9, 11]과 동일하게 본 논문에서도 백채널의 유형 뿐만 아니라 NoBC를 추가하였다. NoBC는 백채널이 발생하지 않음을 나타내며 [11]에서 제안한 방법과 동일하게 백채널 발생 시점으로부터 2초 전을 NoBC로 부착하였다. 이때 NoBC와 백채널이 겹치는 경우에 NoBC 데이터는 제거하였다.

표 1. 한국어 심리 상담 데이터의 백채널 통계

Category	NoBC	Continuer (Cont.)	Understanding (Und.)	Empathetic (Emp.)
# of data	8,513 (41.9%)	9,676 (47.6%)	1,328 (6.5%)	805 (4%)

총 20,322개의 NoBC를 포함한 백채널 데이터를 수집하였으며 각 유형별 데이터 통계는 표 1과 같다. 실험은 5개의 교차검증으로 진행되었으며 데이터를 학습, 검증 및 평가 셋으로 나누는 비율은 3:1:1로 설정하였다.

4.2 하이퍼 파라미터

표 2는 실험에서 사용된 백채널 예측 모델의 하이퍼 파라미터이다. 발화 텍스트를 임베딩하기 위한 BERT 모델은 어절 단위의 한국어 BERT[1]를 사용하였다. 또한 음성 신호를 처리하기 위해 wav2vec 2.0[2]을 사용하였다. 발화 텍스트의 임베딩과

[1]https://aiopen.etri.re.kr/
[2]https://huggingface.co/facebook/wav2vec2-base

표 2. 모델 하이퍼 파라미터

Hyper-parameters		Value
Hidden Dimension Size	BERT	768
	wav2vec	768
	Classifier	1024,256,64
Learning Rate	Pre-trained	0.0001
	Others	0.0007
Warm-up	Pre-trained	0.3
	Others	0.1
Optimizer		AdamW
Weight Decay		0.001
Dropout Rate		0.1

음성 신호 임베딩의 차원 수는 모두 768이다. 분류기는 4개의 선형 계층으로 구성되어 있으며 각 계층의 차원 수는 1024, 256, 64이다. 학습과 관련된 하이퍼 파라미터는 사전 학습 모델(BERT, wav2vec)과 그 외의 모듈을 구분하여 설정하였다. 학습률은 각각 0.0001과 0.0007으로 설정하였고 학습 스케쥴러의 warm-up은 0.3과 0.1을 사용하였다. 최적화 함수는 AdamW를 사용하였으며 weight decay를 0.1로 설정하였다. 또한 드롭아웃은 0.1을 사용하였다. 화자 임베딩은 상담자와 내담자로 구분하여 사용하였다.

4.3 실험 결과

한국어 심리 상담 데이터는 백채널 유형별로 데이터 개수에 상당한 차이가 있다. 특히 Understanding과 Empathetic의 데이터가 전체 데이터의 약 10%만을 차지하고 있다. 불균형한 데이터으로 인해 이전 연구에서 사용한 Weighted-F1 대신에 Macro-F1(M-F1)을 사용하였다. 또한 각 유형별로 F1도 함께 비교하였다.

4.3.1 메모리 크기

표 3은 메모리 크기에 따른 실험 결과이다. 실험 결과로는 메모리가 커질수록 백채널 예측 성능이 향상됨을 볼 수 있었다. 구체적으로 메모리를 7개 사용하였을 때 M-F1 점수가 39.53으로 가장 좋은 성능을 보였다. M-F1 뿐만 아니라 백채널 유형별 F1 점수도 메모리 7개를 사용할 경우에 모든 유형에서 성능이 가장 좋았다. 이는 더 많은 이전 발화의 정보를 사용함으로써 현재 발화의 내용을 더욱 잘 이해하고 적절한 백채널을 예측한 것으로 볼 수 있다.

4.3.2 대화의 맥락 요약 방법

본 논문에서는 대화의 맥락을 요약하는 방법으로 전체 요약과 선택 요약을 제안하였다. 표 4는 대화의 맥락을 요약하는 방법에 따른 성능 비교표이다.

표 3. 메모리 크기별 성능 비교

Memory Size (k)	M-F1	NoBC	Cont.	Und.	Emp.
3	37.16	68.58	63.24	12.32	4.52
5	38.13	69.68	65.63	14.44	2.77
7	**39.53**	**69.98**	**66.81**	**14.79**	**6.55**

표 4. 대화의 맥락 요약 방법에 따른 성능 비교

	C_{HOL}	C_{SEL}	M-F1	NoBC	Cont.	Und.	Emp.
1	-	-	37.71	69.26	65.25	13.13	3.22
2	+	-	38.32	68.92	62.77	**15.34**	6.23
3	-	+	37.44	68.85	61.53	13.27	6.09
4	+	+	**39.53**	**69.98**	**66.81**	14.79	**6.55**

먼저 이전 발화 정보 사용 유무(row1 vs row4)에 따른 성능을 비교하면 발화 정보를 모두 사용하였을 경우에 M-F1이 높은 것을 볼 수 있다. 특히 Understanding과 Empathetic에서 성능이 개선되었다. Continuer는 내용과 관계없이 화자의 발화를 유도하는 역할을 하지만 Understanding과 Empathetic은 화자의 말에 이해하고 공감의 표현을 하는 것이다. 그렇기 때문에 대화의 맥락을 요약하여 사용함으로써 화자의 발화를 이해하고 적절한 백채널을 예측한 것으로 보인다.

전체 요약(C_{HOL})과 선택 요약(C_{SEL})을 각각 적용하였을 때 M-F1 점수는 각각 38.32와 37.44으로 전체 요약의 영향력이 강한 것을 볼 수 있다. 선택 요약만 사용한다면 NoBC와 Continuer의 F1 점수가 크게 낮아져 가장 낮은 M-F1 성능을 보였다. 그렇지만 전체 요약과 선택 요약을 동시에 사용할때 39.53으로 가장 좋은 성능을 보였다.

4.3.3 이전 연구와 비교

본 논문에서 제안한 모델과 이전의 백채널 예측 모델을 비교한다. Ortega[8]는 CNN으로 고정된 크기의 텍스트와 음성 신호를 인코딩하여 백채널을 예측하였다. [9]는 고정된 길이의 텍스트를 BERT에 입력으로 넣어주어 출력된 임베딩을 음성 신호 임베딩과 결합하여 백채널 예측을 수행하였다(BPM_ST). 추가로 상담 대화에서 백채널 예측 성능을 향상시키기 위해 감성 분류 작업을 포함하여 다중 학습 방식으로 훈련시켰다(BPM_MT). 비교를 위해 본 논문에서 제시한 데이터를 활용해 재구현하고 평가하였다.

표 5은 이전 연구에서 제안된 모델과 성능을 비교한 표이다. Ortega는 30.85의 M-F1 점수를 받았으며 BPM_ST와

표 5. 이전 연구와의 성능 비교(†: 재구현)

Model	M-F1	NoBC	Cont.	Und.	Emp.
Ortega†[8]	30.85	59.52	59.28	1.02	3.58
BPM_ST†[9]	34.47	62.60	59.82	11.06	4.39
BPM_MT†[9]	35.00	61.30	59.23	**16.69**	4.77
Ours	**39.53**	**69.98**	**66.81**	14.79	**6.55**

BPM_MT는 각각 34.47과 35의 성능을 보였다. 마지막으로 본 논문에서 제안한 모델의 M-F1 점수는 39.5으로 가장 좋은 성능을 보이고 있다. Understanding을 제외한 백채널에서 성능 차이가 크게 나타나고 있다.

5. 결론

벡채널은 대화 상황에서 화자가 말을 계속 할 수 있도록 장려하고 공감을 표현하는 방법 중 하나이다. 이때 적절한 백채널을 예측하기 위해서는 현재 발화의 의미를 파악하는 것이 중요하며 이것은 대화의 맥락을 이해하는 것이 필요하다. 본 논문에서는 대화 맥락을 요약하는 방법을 제안하여 백채널 예측의 성능을 개선하고자 한다. 대화 맥락을 요약하는 방법으로 대화 전체를 요약하는 전체 요약과 현재 발화에서 생략된 정보를 보완하는 선택 요약을 제안하였다.

한국어 심리 상담 데이터로부터 실험 결과는 현재 발화만 사용하였을 때보다 이전 발화를 통해 대화 맥락을 반영하였을 경우에 백채널 예측 성능이 개선됨을 보였다. 특히 화자의 발화를 이해하고 공감하는 백채널인 Understanding과 Empathetic에서 성능이 개선되었다.

감사의 글

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2022-0-00608, 인간과 교감하는 멀티모달 인터랙션 인공지능 기술)

참고문헌

[1] L. Tickle-Degnen and R. Rosenthal, "The nature of rapport and its nonverbal correlates," *Psychological inquiry*, Vol. 1, No. 4, pp. 285–293, 1990.

[2] V. H. Yngve, "On getting a word in edgewise," *Papers from the sixth regional meeting Chicago Linguistic Society, April 16-18, 1970, Chicago Linguistic Society, Chicago*, pp. 567–578, 1970.

[3] L. Huang, L.-P. Morency, and J. Gratch, "Virtual rapport 2.0," *Intelligent Virtual Agents*, H. H. Vilhjálmsson, S. Kopp, S. Marsella, and K. R. Thórisson, Eds., pp. 68–79, 2011.

[4] B. Xiao, P. G. Georgiou, Z. E. Imel, D. C. Atkins, and S. S. Narayanan, "Modeling therapist empathy and vocal entrainment in drug addiction counseling." *Interspeech*, pp. 2861–2865, 2013.

[5] C. Goodwin, "Between and within: Alternative sequential treatments of continuers and assessments," *Human studies*, Vol. 9, No. 2-3, pp. 205–217, 1986.

[6] P. Blache, M. Abderrahmane, S. Rauzy, and R. Bertrand, "An integrated model for predicting backchannel feedbacks," *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, pp. 1–3, 2020.

[7] J. Y. Jang, J. Kim, M. Jung, H. Jung, and S. Shin, "Utilizing multi-modal emotion information in dialogue strategy classification," *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 943–945, 2020.

[8] D. Ortega, C.-Y. Li, and N. T. Vu, "OH, JEEZ! or UH-HUH? a listener-aware backchannel predictor on ASR transcriptions," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8064–8068, 2020.

[9] J. Y. Jang, S. Kim, M. Jung, S. Shin, and G. Gweon, "BPM_MT: Enhanced backchannel prediction model using multi-task learning," *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 3447–3452, Nov. 2021. [Online]. Available: https://aclanthology.org/2021.emnlp-main.277

[10] A. I. Adiba, T. Homma, and T. Miyoshi, "Towards immediate backchannel generation using attention-based early prediction model," *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7408–7412, 2021.

[11] R. Ruede, M. Müller, S. Stüker, and A. Waibel, "Enhancing Backchannel Prediction Using Word Embeddings," *Interspeech 2017*, pp. 879–883, 2017.

[12] K. Hara, K. Inoue, K. Takanashi, and T. Kawahara, "Prediction of turn-taking using multitask learning with prediction of backchannels and fillers," *Interspeech 2018*, pp. 991–995, 2018.

[13] G. Malhotra, A. Waheed, A. Srivastava, M. S. Akhtar, and T. Chakraborty, "Speaker and time-aware joint contextual learning for dialogue-act classification in counselling conversations," *Proceedings of the 15th ACM International Conference on Web Search and Data Mining*, pp. 735–745, 2022.

[14] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *Advances in neural information processing systems*, Vol. 33, pp. 12 449–12 460, 2020.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Jun. 2019. [Online]. Available: https://aclanthology.org/N19-1423

[16] E. Ribeiro, R. Ribeiro, and D. M. de Matos, "The influence of context on dialogue act recognition," *arXiv preprint arXiv:1506.00839*, 2015.

[17] L. Meng and M. Huang, "Dialogue intent classification with long short-term memory networks," *Natural Language Processing and Chinese Computing: 6th CCF International Conference, NLPCC 2017, Dalian, China, November 8–12, 2017, Proceedings 6*, pp. 42–50, 2018.

[18] Z. Pan, K. Bai, Y. Wang, L. Zhou, and X. Liu, "Improving open-domain dialogue systems via multi-turn incomplete utterance restoration," *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 1824–1833, 2019.

[19] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.