

음성 기반 상담의 품질 평가를 위한 자동화 기법[☆]

A Method of Automated Quality Evaluation for Voice-Based Consultation

이 건 수¹ 김 중 연^{2*}
Keonsoo Lee Jung-Yeon Kim

요 약

언택트 시대의 시작으로, 온라인 산업의 성장 속도는 점차 빨라지고 있다. 온라인 산업이 성장할수록, 고객 관리에 대한 중요성은 높아지며, 그 점점에 존재하는 컨택센터 시장 역시 성장하고 있다. 언택트 시대의 주요 서비스 분야인 컨택센터의 업무가 노동 집약적이라는 아이러니를 극복하고 컨택센터 업무 효율을 증가시키기 위한 다양한 업무 자동화 기술 개발 연구들이 활발하게 진행되고 있다. 본 연구는 업무 자체는 정형적이지만, 그 중요성이 높아 업무 자동화의 효율이 높은 대표적인 컨택센터의 업무 중 하나인 품질 평가 업무의 자동화 방법을 제안한다. 제안 방법은 채널 분리된 상담 내용 녹취 내용을 토대로, 음성 인식 결과를 획득한 뒤, 문장 단위 발화 내용을 분석하여, 정량 평가 항목인 도입부 평가, 응대 중 경청과 침묵 평가, 그리고 마무리 평가를 수행한 후 수행 결과를 평가표에 맞춰 출력하는 단계를 따른다. 제안 방법은 전문가의 평가 결과 대비 92.7% 일치율을 보였다. 불일치 케이스의 경우, 주로 음성 인식의 오류에 기인한 경우였다. 따라서 음성 인식 결과의 신뢰도가 보장된다면, 본 논문에서 제안한 방법을 통해 자동화된 품질 평가로 해당 업무 효율을 증대시킬 수 있을 것이다.

☞ 주제어 : 컨택센터, 상담, 품질평가, 음성 인식, 업무 자동화

ABSTRACT

In a contact-free society, online services are becoming more important than classic offline services. At the same time, the role of a contact center, which executes customer relation management (CRM), is increasingly essential. For supporting the CRM tasks and their effectiveness, techniques of process automation need to be applied. Quality assurance (QA) is one of the time and resource consuming, and typical processes that are suitable for automation. In this paper, a method of automatic quality evaluation for voice based consultations is proposed. Firstly, the speech in consultations is transformed into a text by speech recognition. Then quantitative evaluation based on the QA metrics, including checking the elements in opening and closing mention, the existence of asking the mandatory information, the attitude of listening and speaking, is executed. 92.7% of the automated evaluations are the same to the result done by human experts. It was found that the non matching cases of the automated evaluations were mainly caused from the mistranslated Speech-to-Text (STT) result. With the confidence of STT result, this proposed method can be employed for enhancing the efficiency of QA process in contact centers.

☞ keyword : Call Center, Consultation, Quality Assurance, Speech Recognition, Process Automation

1. 서 론

2019년 12월 중국에서 처음 보고된 코로나바이러스감염증-19는 1년이 넘는 현재까지 전 세계에서 사람들을 감염시키며, 인류의 삶의 방식을 변화시켰다. 소위 언택트 시대라고 불리는 새로운 삶의 모습은 사람들이 직접 만

나는 것 대신 ICT 기술을 통한 원거리 통신을 기본적인 사회적 의사소통 방식으로 정착시켰다[1].

온라인 쇼핑 거래액의 경우 2020년 11월 기준으로 2019년 대비 17.2%가 증가한 15조 631억원을 기록했으며, 오프라인 쇼핑의 내림세와 함께 온라인 쇼핑의 상승세는 지속될 전망이다^{*}. 이와 함께 비대면 의사소통의 주요 수단인 전화 기반의 서비스들에 대한 수요는 증가하고 있다. 1339 콜센터의 경우, 업무량이 10배 증가하였으며, 동시에 컨택센터 특유의 밀집된 근무 형태에 대한

¹ Research Lab, CNAI Ltd., Seoul, 06711, Korea.

² ICT Convergence Research Center, Soonchunhyang University, Asan, 31538, Korea.

* Corresponding author (betterwithme@sch.ac.kr)

[Received 28 January 2021, Reviewed 9 February 2021, Accepted 2 March 2021]

[☆] 이 연구는 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2020R1I1A1A01073408).

[†] https://kosis.kr/common/meta_onedepth.jsp?vwcd=MT_ZTITLE&listid=O_14

[‡] https://www.ytn.co.kr/_ln/0103_202011070029348193

대안으로 원격 근무의 요구 또한 증가하고 있다*.

이러한 산업 환경 변화는 업무 효율 개선을 통해 대응할 수 있는데, 소위 ‘인공지능 컨택센터’라고 지칭되는 컨택센터 자동화 기술의 도입이 그 수단이 된다. 컨택센터 업무 중 자동화 대상 업무는 크게, 상담사의 상담 업무와 상담사 관리 업무로 나눌 수 있다. 전자의 경우 챗봇과 상담 지원 솔루션이 주로 사용되고 있지만, 후자의 경우 특별한 표준 기술 없이 각 세부 업무에 대한 산업공학적 분석을 통한 효율 개선 방법을 적용하고 있다[2].

대표적인 상담사 관리 업무로는 상담 품질 평가가 있다. 이 과정은 다음의 흐름을 따르게 되는데, 우선 상담원이 고객과 통화한 내용은 녹취된다. 이후 품질평가 과정을 거쳐 상담사의 상담 능력 향상 및 추가 교육을 위한 자료로 활용된다. 하지만 일반적으로, 컨택센터에서 상담사의 수는 품질 평가를 수행하는 교육 강사의 수를 상회하기 때문에 모든 상담사의 통화 내용을 검수하는 것은 불가능하다. 이러한 이유로 인해 특정 상담사의 특정 상담 내용에 대한 평가를 수행하거나 무작위 선별을 통한 샘플 상담 내용을 토대로 평가를 수행하는 방법을 사용하게 되는데, 우연에 근거한 상담 업무 평가이다 보니, 이를 토대로 인센티브 등을 통한 인적자원관리에 유의미한 근거 자료로 활용하기 힘들다.

이에 본 연구에서는 음성 기반의 상담 녹취 데이터를 대상으로 상담 품질 평가를 자동으로 수행하는 방법을 제안하였으며, 실제 구현을 통해 자동화된 평가 결과와 수작업으로 진행된 평가 결과를 비교하였다.

2. 연구 배경

2.1 품질 평가 방법

품질 보증(QA : Quality Assurance)은 제품 생산에서부터 납품, 그리고 납품된 제품의 운용과정에서 발생하는 불량률을 낮추기 위한 방법으로 사용되고 있다. ISO 9000* 에서는 “품질이 요구사항을 만족시키고 있다는 확신”을 만드는 과정이라고 정의하고 있다. 현재 일반적으로 사용되는 QA기법으로 PDCA[3], DMAIC[4], DFSS[5], RADAR[6] 등이 존재한다. PDCA의 경우, 품질 정책을 생성하고 적용하기 위해 사용된다. DMAIC과 DFSS의 경우, 제품의 생명주기에 정밀성을 보장하기 위해 사용되며,

RADAR의 경우, 개별 제품보다는 조직 자체의 탁월성을 향상시키기 위한 방법으로 사용된다. 앞서 나열된 품질 보증 방법들은 질적 향상의 대상이 다를 뿐, 그 방법 자체는 서로 유사하다.

우선 PDCA의 경우, Plan, Do, Check, 그리고 Act의 과정으로 구성된 사이클을 반복한다. 계획을 세워서 적용하고, 그 결과를 확인해서 효과가 있으면 반영하는 과정을 반복해서 적용하는 방법이다. DMAIC은 Define, Measure, Analysis, Improve, 그리고 Control의 과정으로 이루어지며, 6-Sigma로 알려진 방법론의 핵심 프로세스다. 목표를 정의하고, 평가를 위한 기준을 생성한 후 기준에 의해 현재 상태를 분석한다. 분석 결과를 바탕으로 개선점을 찾아 개선하고, 개선 결과를 유지하는 것으로 프로젝트를 완료한다. Design for Six Sigma (DFSS)는 6-Sigma를 위한 디자인 방법론으로써 실행착오를 줄이기 위해 처음 설계과정에서 최선의 결과를 도출하기 위한 방법이다. DFSS를 적용하기 위한 방법으로 DMADV와 IDOV 방법을 선택할 수 있는데, DMADV는 Define, Measure, Analyze, Design, 그리고 Verify로 이루어진 과정인데, 이는 DMAIC과 거의 유사하다. DMAIC가 제품의 개선과 유지를 위함이라면, DMADV는 제품을 위한 디자인과 그 검증이라고 할 수 있다. 반면 IDOV는 Identify, Design, Optimize 및 Validate의 과정으로 이루어진다. DMADV와 IDOV의 가장 큰 차이라면, 최종단계가 디자인에 대한 검사(Verification), 또는 검증(Validation)을 수행하는지에 달려있다고 할 수 있다. RADAR의 경우, Result, Approach, Deploy, Assess, 그리고 Refine의 다섯 요소를 기반으로 하는 사이클을 반복하는 방법으로 목표 결과를 달성하기 수단을 계획하고, 적용한 후, 그 적용 결과를 평가하여 개선하는 과정을 반복한다. 상담 결과에 대한 QA는 개별 상담사가 각 고객에 대해 제공하는 서비스가 요구사항을 만족하는 것인지를 판단하는 과정인데, 품질 보증 전체 과정을 아우르는 것이 아닌, 평가 과정을 의미한다. 즉, DFSS와 RADAR 보다는 PDCA 혹은 DMAIC의 적용 대상이며, PDCA의 Check 또는 DMAIC의 Analysis 단계에서 수행해야 하는 업무라고 할 수 있다.

2.2 상담 품질 평가를 위한 기준

상담 내용의 품질 평가 기준은 각 컨택센터가 담당하고 있는 업무 내용 및 센터의 운영 철학에 따라 조금씩 상이한 모습을 보이지만, 공통적으로 다음의 다섯가지 요소로 분류 가능하다(표 1).

* <https://www.bbc.com/korean/features-51867202>

† <https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en>

(표 1) 일반적인 상담 QA 평가 기준표
(Table 1) General metrics for quality evaluation on consultations

요소	내용
도입	상담을 시작할 때, 인사말, 소속, 이름으로 구성된 첫인사를 발화해야 한다.
응대	응대 과정에서 상담원은 1) 고객의 말을 경청하고, 2) 목음 상태를 유지하면 안되고, 3) 상담시 필요한 고객 정보를 확인해야 하며, 4) 고객의 문의에 정확한 정보를 제공해야 하고, 5) 제공 과정에서 고객이 이해하기 쉽도록 설명해야 한다.
마무리	상담 종료시, 추가문의여부를 확인하여야 하며, 더 이상의 추가 문의가 없을 경우, 소속, 이름, 종료멘트로 구성되는 끝인사를 발화해야 한다.
가점항목	상담을 친절하게 처리했거나, 상담 중 쿠션어를 n회 이상 사용했을 때 가점된다.
감점항목	불친절하게 상담을 진행하거나, 상담 결과 이력 기록 과정이 미흡했을 때 감점된다.

이들 평가 기준 중에서, 정성 평가인 경우는 자동화 처리가 용이하지 않으며, 이해당사자들 사이의 이견 조율이 필요하기에, 본 연구의 자동화 대상에서 제외하였다. 가령, 정확도, 설명력, 불친절 항목의 경우, 평가자에 따라서 다르게 평가될 수 있는 여지가 있으며, 점수 배점에 대한 명확한 기준이 존재하지 않았다. 설명력 항목의 만점 기준은 “고객의 수준에 맞춰 전문용어를 쉽고 자세히 설명하여 이해도를 높이는 경우”인데, 이처럼, 상담사의 발화 뿐만 아니라, 고객과의 상호 작용을 추적해서, 각각의 고객에 맞는 설명이 이루어졌는가를 판단하기 위해서는 보다 많은 상황 정보가 요구된다. 반면 정량 평가의 대상이 되는 요소들은 이견의 여지가 존재하지 않으므로 평가 규칙이 정해지면 상대적으로 쉽게 자동화시킬 수 있다. 본 연구에서 선택한 정량 평가 기준은 3.2장에서 기술한다.

3. 제안 방법

3.1 요구사항

22명의 상담직원이 근무하고 있는 A컨택센터의 경우, 2020년 10월 20일 기준으로 1,131건의 상담 요청 전화가 왔고, 그 중 951건의 전화가 연결되었다. 평균 상담시간은 3분 47초였지만, 전체의 2분 미만의 상담이 25%이고, 3분 미만의 전화는 50%였으며 8분을 초과하는 전화가 전체의 10%를 차지하고 있었다. 이 컨택센터의 경우, 주요

상담 업무가 신규계약, 정산일 및 금액 확인, 신용카드 거래내역 확인 등이기 때문에, 8분을 넘어가는 상담은 대부분이 특이 상담으로 민원 및 고객 불만으로 재접수 된다. 이런 예외 상담을 제외하면, 전체 상담의 90% 정도를 포함하는 8분 미만의 상담 전화에 대해서는 집중적이고 예외적인 품질 검증이 아닌, 일반적인 검사로 품질 검증이 가능하다. 하지만 품질에 문제가 없을 가능성이 높은, 즉 고객의 민원으로 연결되지 않는 일반적인 상담의 경우, 그 상담의 수가 많아서 소수의 평가원이 수작업으로 전수 검사하는 것이 불가능하다. 22명의 상담원을 보유하고 있는 해당 센터에서도 검수는 2명의 교육 강사에 의해 처리되기에, 전수 검사보다는 소수의 샘플을 선택하여 검사하는 방식을 사용한다.

품질 검사는 2.2장에서 기술된 검사 기준 중 정량 평가가 가능한 요소를 대상으로 수행한다. 평가 결과는 JSON 형식으로 반환하며, 이 결과는 기존의 UI와 결합해 사용자에게 제공되며, 동시에 외부 UI 시스템과 결합되지 않더라도 사용자가 결과를 확인할 수 있도록 결과의 문서화 기능이 포함된다.

3.2 필요기능

3.2.1 도입부 - 인사말 평가 기능

상담이 시작되는 도입부에서 인사말은 ‘인사’, ‘소속’, ‘이름’의 정보로 구성된다. 대화가 시작되고 4초 이내에 이 정보들이 발화되었는지에 대한 여부로 판단한다. 평균적으로 이러한 정보의 발화는 2초 내외의 시간을 소요하는데, 평균 발화 시간의 두 배를 발화 시간 범위로 정했다. 해당 시간 범위에서 발화된 음성을 문자로 변환한 뒤, 형태소 분석 및 개체명인식(NER: Named Entity Recognition)을 사용해, 각각의 요소가 발화 안에 존재하는지를 확인한다 [7]. 인사는 다양한 변이를 허용하기 위해서, 형태소 분석 결과 중 핵심 공통 키워드가 존재하는 경우 인사라고 판단한다. 소속은 정해진 소속이 정확하게 존재하는지를 판단하며, 이름의 경우 NER을 통해 사람인 경우를 찾는다. NER 작업은 Transformer 기반의 BERT[8] 혹은 ELECTRA[9] 등의 모델을 통해 높은 수준의 결과물을 도출할 수 있지만, 사전 학습 및 미세조정 등의 준비 과정들이 추가로 필요하다. 또한 일반 명사와 유사한 이름의 경우 오처리 되는 경우가 많으므로 소규모 컨택센터의 경우, 전체 상담사 명을 직접 비교하는 방식을 적절하게 혼용한다.

3.2.2 응대부 - 경청능력 평가 기능

경청능력은 상담사가 고객의 발화 내용을 얼마나 주의 깊게 들었는지에 대한 정량적인 평가 요소로, 고객의 발화 중 상담사가 말을 자르고 개입한 경우를 탐지해 그 횟수로 판단한다. 이때 추입새나, 감탄사의 경우를 제외하기 위해서, 다음과 같은 휴리스틱 규칙을 추가한다. 첫째, 발화의 겹침은 항상 고객의 발화가 먼저 진행 중인 상태에서 상담사의 발화가 겹치는 경우여야 한다. 둘째, 겹침이 발생했을 때, 0.5 초 이상의 겹침이 발생해야 하고, 겹침 이후, 상담사의 발화가 연속해서 발생해야 한다. 셋째, 겹침 이후 고객이 말을 이어서 했다고 해도, 겹침의 구간이 1초를 초과하는 경우, 개입이 발생한 것으로 판단한다.

3.2.3 응대부 - 묵음 평가 기능

상담 과정에서 고객의 요구를 충족시키기 위해서, 외부 정보를 참조해야 하는 경우 발생하는 묵음은 허용할 수 있지만, 그 이외의 경우 발생하는 묵음은 상담 품질을 낮추는 근거가 된다. 대기 요청이 없는 상태로 발생하는 묵음 구간을 탐지하기 위해 다음의 과정을 따른다. 우선 5초 이상 진행되는 묵음 구간을 탐지하고, 해당 묵음의 시작 바로 전의 상담사의 발화를 탐지한다. 만약, 상담사의 발화에 대기 요청 문구가 포함되어 있다면, 해당 묵음은 예외로 처리하고, 그렇지 않은 경우는 묵음 발생 구간으로 판단한다.

3.2.4 응대부 - 정보확인 평가 기능

정보확인 은 고객이 요청하는 서비스를 제공하기 위해, 고객이 서비스의 접근 권한이 있는지를 확인하기 위한 과정으로, 고객이름, 고객연락처는 공통으로 확인해야 하는 부분이며, 이후, 상담 내용에 따라, 계좌번호, 주문번호, 상호명 등을 포함한다. 상담사는 고객에게 해당 내용을 요청하고, 고객의 발화 내용을 복창하는 것으로 정보확인 절차가 끝이 난다. 이때 확인해야 하는 내용들이 개인정보에 해당하기 때문에, 이 부분은 정보확인 절차가 진행되었는지만을 판단한다. 상담사는 고객의 이름과 연락처를 요청하고, 고객으로부터 전달받은 내용을 복창하며, 마지막으로 '확인 감사합니다'라는 발화를 수행하게 되는데, 이 절차가 상담 내용에 존재하는지를 판단하는 것으로 해당 평가 기능을 갈음한다.

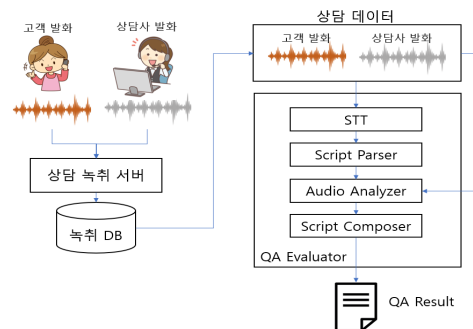
3.2.5 마무리 - 끝인사 평가 기능

상담이 종료되는 마무리 부분에서 끝인사는 '추가문의 여부', '소속', '이름', '끝인사'의 정보로 구성된다. '추가문의여부'로는 "다른 문의사항은 없으십니까?" 와 같은 질문으로 '[추가가다른디] 문의사항은 [있없][으십니까나 오으세요?]' 같은 표현으로 변이 가능하다. 이후 고객이 이 질문에 긍정하면 마무리 단계는 다시 응대 단계로 되 돌아간다. 고객의 부정 확인 이후, 도입부에서 수행했던 내용과 동일한 과정을 진행한다. 다만, 인사말은 '고맙습니다'나 '감사합니다'가 된다. 끝인사는 상담원의 마지막 말이 되기에, 상담사의 발화 중 마지막 5문장 안에서 마무리 인사를 탐색한다. '소속'과 '이름'을 탐색하는 과정은 도입부에서 사용했던 흐름과 동일하다.

4. 구현 결과

4.1 상담 품질 자동 평가 어플리케이션

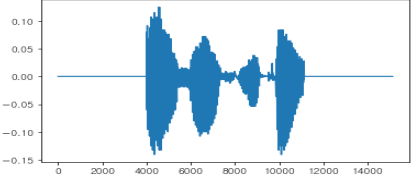
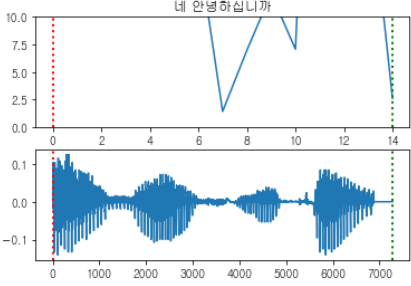
제안 방법은 다음과 같은 구조로 구현되었다. 상담사의 상담 내용은 상담사의 발화와 고객의 발화로 구분되어서 녹취된다. 상담 내용은 녹취 데이터베이스에 저장되어 있다가, 상담이 끝난 후에, 상담 내용에 대한 품질 평가 과정이 적용된다.



(그림 1) 제안 솔루션의 아키텍처 구조
(Figure 1) System Architecture of the proposed application

본 어플리케이션은 4개의 단계를 통해, 최종 QA 평가 결과를 도출한다. 각각의 단계는 해당 단계를 책임지는 모듈을 통해 구현된다. 다음 표는 각 단계에서 입력으로 받는 데이터와 해당 단계의 처리를 통해 결과로 도출되는 결과를 보여준다.

(표 2) 제안 방법의 단계별 입력 및 출력 결과
(Table 2) Input and Output of each process step in the proposed process

입력 데이터 (녹취 파일)	
Step 1 : STT	
출력 결과	[['네', 530.0, 680.0], ['안녕하십니까', 680.0, 1440.0]]
Step 2 : Script Parser	
출력 결과	[['sentence': '네 안녕하십니까', 'start': 530.0, 'end': 1440.0, 'number': 0]]
Step 3 : Audio Analyzer	
출력 결과	 [['start': 530, 'end': 614, 'sentence': '네 안녕하십니까']]
Step 4 : Script Composer	
출력 결과	<pre> { "Result": { "Greeting": { "Greeting": True/False, "Affiliation": True/False, "Name": True/False }, "Interruption": { "Num": count, "Positions": [{"start": number, "end": number}, ...] }, "Mute": { "Num": count, "Positions": [{"start": number, "end": number}, ...] }, "InfoCheck": { "Completion": True/False }, "Greeting": { "Addition": True/False, "Affiliation": True/False, "Name": True/False, "Greeting": True/False } } } </pre>

첫 번째 단계를 담당하는 Speech-to-Text (STT)는 선택한 상담 녹취 파일을 기반으로 상담사와 고객의 발화 내용을 각각 텍스트로 변경한다. STT 엔진은 (주)CnAI의 음성인식 엔진을 사용하였다. 본 구조에서의 STT 모듈은 녹취 서버로부터 획득한 음성 파일을 음성인식 엔진에 전달하고, 그 결과를 받아와 다음 단계의 모듈에게 전해주는 역할을 수행한다. STT 단계의 결과물은 음성의 인식 결과와 해당 결과의 시간 정보의 집합이다.

두 번째 단계를 담당하는 Script Parser는 STT의 결과물을 토대로, 발화의 스크립트 문장을 생성한다. 상담 도입 부분에서의 첫인사 평가, 응대 과정에서의 말 끊기, 무음 구간, 정보확인, 그리고 마무리 과정에서 끝인사에 대한 발화를 평가하기 위해서는 문장 단위로 처리하는 것이 유용하기 때문에, STT의 결과를 문장 단위의 스크립트로 변경한다.

세 번째 단계를 담당하는 Audio Analyzer는 이전 단계에서 구분된 문장을 음성 정보를 토대로 검증한다. 이는 상담 과정에서 상담사와 고객이 완전한 문장으로 대화를 주고받는 것이 아니라, 불완전한 문장으로 대화를 주고받을 수 있지만, STT의 결과만으로는 이런 비정형 문장 구분을 인지할 수 없다. Audio Analyzer는 Mel Spectrogram을 이용해, 발화의 무음 구간을 추출하고, 이를 기준으로 문장을 구분한다[10]. 가령, “아 그건 말이죠 (공백) 이만 오천원입니다” 라는 발화의 경우, 하나의 문장이 아니라 두 개로 분리 되어야 하기 때문이다.

마지막 단계를 담당하는 Script Composer는 분리된 문장 목록을 대상으로 인사말 평가, 경청능력 평가, 묵음평가, 정보확인 평가, 끝인사 평가를 각각 수행한 뒤, 그 결과를 정의된 형식으로 결합하여 출력한다. 평가 방법은 앞서 3.2장에서 기술된 흐름을 따른다.

4.2 결과 유효성 검증

제안 방법을 통해 구현된 어플리케이션의 결과를 검증하기 위해서, 3.1장의 컨택센터로부터 30개의 QA 평가서와 개인정보를 제외한 상담 데이터를 제공받았다.

(표 3) 실험 환경

(Table 3) Machine Specification for the experiment

CPU	Intel Xeon® 4114 2.2GHz
RAM	64G
VGA	Geforce RTX 2080 ti (2대)
OS	CentOS version 7

개인 정보는 정보확인평가를 위해 고객이 발화한 연락처와 이름 중 연락처 정보를 제거한 후 해당 시간대의 음성 정보를 해당 음성의 평균 에너지 값으로 대체하는 식으로 처리하였다. 이 과정에서, 평균 에너지가 검출되는 구간을 연락처 정보가 위치하는 구간이라고 판단해버리면, 정보 확인 평가를 올바르게 수행할 수 없기 때문에, 해당 검증 과정에서 정보 확인은 이름을 확인한 것으로

축소해서 처리하였다. 실험을 수행하기 위해 사용한 컴퓨터의 성능은 (표 3)과 같다.

제안 방법 검증을 위해 사용된 상담 녹취 파일은 2020년 12월 09일에서 2020년 12월 18일 사이에 발생한 상담 중 11명의 상담사에 의해 처리된 상담 녹취 데이터에서 30개를 무작위로 선택하였다. 휴일을 제외한 8일 동안의 녹취된 상담은 총 5,472건으로 하루 평균 684개의 상담이 진행되었다. 이들 중 30개의 선택은 실제 상담 내용의 QA를 진행하는 센터의 강사들이 QA 업무를 수행하기 위해 선택된 것으로 센터 운영진의 QA 결과와 얼마나 유사하게 자동으로 QA를 처리할 수 있는지를 비교하였다.

전체 데이터의 경우, 4분 37초 미만의 전화가 전체의 71.5%를 차지하고 있었고, 7분 미만이 16.9%, 9분 미만이 6.3% 그리고 그 이후가 4.8%의 비율을 구성하고 있었다. QA 담당자가 선택한 녹취는 13개의 4분 구간, 9개의 5분 구간, 6개의 6분 구간, 그리고 2개의 7분 구간의 상담들로 구성되었다. 검증을 위해 무작위로 선택된 녹취 샘플들은 실제 녹취 데이터의 분포와 비슷한 경향을 보이고 있었으며, 10분이 넘어가는 상담 대부분의 경우, 고객의 불만으로 인해 민원 업무로 이관되기 때문에, QA의 범위에 포함되지 않아 검증 녹취 샘플에 포함되지 않았다.

(표 4) 실험 결과

(Table 4) Experimental Result

	일치	불일치
인사말평가	28	2
기준	수작업 결과와 동일해야 함	
경청능력평가	30	0
기준	수작업 결과를 모두 포함하면 일치	
목음평가	30	0
기준	수작업 결과를 모두 포함하면 일치	
정보확인평가	30	0
기준	해당 과정이 존재하지 않는 경우를 판별했다면 일치	
끝인사평가	21	9
기준	수작업 결과와 동일해야 함	
총수	139	11

QA전문가의 평가 결과와 제안 방법을 이용한 평가 결과를 비교하였다. 30개의 상담 녹취에 대해서, 각각 5개의 요소들을 평가하기 때문에, 총 150개의 요소에 대한 결과를 비교하였다. 비교 기준 및 결과는 다음과 같다 (표4).

인사말과 끝인사 평가의 경우, 전문가 평가와 동일한 결과를 일치로 판단하고, 경청능력 및 목음 평가의 경우, 전문가 평가를 모두 포함하고 있다면, 일치로 판별하였

다. 이는 평가 기준을 적용함에 있어서, 사람이 귀로 듣고 판단하는 것보다 시스템이 비교하는 것이 더 꼼꼼하게 찾아내기 때문이다. 정보확인평가의 경우, 확인하는 정보의 상세 내용을 구분하지 않기 때문에, 해당 과정의 존재 유무를 판별하는 경우를 일치로 판단하였다.

일치율은 150개 항목 중 139개의 일치로 92.7%의 정확도를 보였다. 불일치한 11개의 경우, 해당 문구의 STT 결과가 올바르게 없어서 불일치한 경우로, 소속이 다르게 인식되거나, 이름이 오인식된 경우, 그리고 예측하지 못한 참신한 인사말을 사용한 경우로 드러났다.

5. 결 론

품질 관리는 제조업뿐만 아니라, 모든 서비스 분야에서도 고객만족도를 향상시키기 위한 필수 요소다. 시간과 비용이 많이 소요되는 노동집약형 업무이면서 동시에 일반적인 업무 특징을 갖고 있기에, 자동화를 통한 업무 효율 향상에 적합하다. 언택트 시대의 흐름에 따라, 업무량이 증가하고 있는 컨택센터의 경우, 자동화를 통한 업무 효율 개선 요소를 많이 갖고 있으며, 특히 인력의 한계로 샘플 검수밖에 하지 못하는 QA 업무의 경우, 본 제안 방법을 통해 전수 검사가 가능하며, 몇 가지 제약점을 제외한다면, 수작업 결과와 동일한 결과를 기대할 수 있다. 본 제안 방법은 상담 음성의 전사 결과를 기반으로 동작하기 때문에, 전사에 오류가 존재하는 경우, 후속 작업의 정확성을 보장할 수 없다. 즉 STT 과정에서 발생하는 오류 발생 시 그에 대한 인지 및 처리에 한계가 존재한다. 또한 QA의 평가 대상이 정량 평가 기준에 한정되어 있다. 정성 평가의 경우, 추후 이해당사자들 사이의 다툼이 발생하거나, 평가 결과의 민감성에 관한 충돌을 피하기 위해, 본 연구에서는 다루지 않았지만, 완전한 QA 업무 자동화를 위해서는 정성 평가 기준에 대한 대응 역시 필요한 부분이다.

이러한 제약을 감수한다면, 본 기술은 원격재활 도메인에서 치료 서비스 제공자가 사용자에게 적절하고 유효한 절차에 따라 치료 서비스를 제공하였는지 파악하거나, 비대면으로 진행되는 인지 기능 평가에서 평가 절차의 준수 여부를 자동으로 확인하기 위한 수단으로 활용될 수 있다.

참고문헌(Reference)

- [1] M. Lee, and D. Lee. ““Untact”: a new customer service strategy in the digital age.” *Service Business* vol. 14. no. 1, pp. 1-22. 2020.
<http://dx.doi.org/10.1007/s11628-019-00408-2>
- [2] M. Rim, K. Thomas, j. Chandramouli, S. Barrus, and N. Nickman. “Implementation and quality assessment of a pharmacy services call center for outpatient pharmacies and specialty pharmacy services in an academic health system”. *The Bulletin of the American Society of Hospital Pharmacists*, vol. 75, no. 10, pp. 633-641, 2018.
<https://doi.org/10.2146/ajhp170319>
- [3] A. Prashar, “Adopting PDCA (Plan-Do-Check-Act) cycle for energy optimization in energy-intensive SMEs”. *Journal of cleaner production*, vol. 145, pp. 277-293, 2017. <https://doi.org/10.1016/j.jclepro.2017.01.068>
- [4] S. Ahmed, “Integrating DMAIC approach of Lean Six Sigma and theory of constraints toward quality improvement in healthcare.” *Reviews on environmental health*, vol. 34, no. 4, pp. 427-434, 2019.
<https://doi.org/10.1515/reveh-2019-0003>
- [5] X. Yang, S. Gao, Z. He, and M. Zhang, “Application of Design for Six Sigma tools in telecom service improvement”. *Production Planning & Control*, vol. 29, no. 12, pp. 959-971, 2018.
<https://doi.org/10.1080/09537287.2018.1486469>
- [6] E. Suárez, A. Calvo-Mora, J. Roldán, and R. Perriñez-Cristóbal, “Quantitative research on the EFQM excellence model: A systematic literature review (1991 - 2015)”. *European Research on Management and Business Economics*, vol. 23, no. 3, pp. 147-156, 2017.
<https://doi.org/10.1016/j.iedeen.2017.05.002>
- [7] J. Li, A. Sun, J. Han, and C. Li, “A survey on deep learning for named entity recognition”. *IEEE Transactions on Knowledge and Data Engineering*. 2020.
<https://doi.org/10.1109/TKDE.2020.2981314>
- [8] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding”. 2018.
 arXiv preprint, <https://arxiv.org/abs/1810.04805>
- [9] K. Clark, M. Luong, Q. Le, and C. Manning, “Electra: Pre-training text encoders as discriminators rather than generators”. 2020.
 arXiv preprint, <https://arxiv.org/abs/2003.10555>
- [10] Graf, S., Herbig, T., Buck, M. et al. “Features for voice activity detection: a comparative analysis”. *EURASIP J. Adv. Signal Process.* vol. 91, 2019.
<https://doi.org/10.1186/s13634-015-0277-z>

◎ 저 자 소 개 ◎



이 건 수(Keonsoo Lee)

2002년 아주대학교 미디어학부(공학사)
 2004년 아주대학교 정보통신전문대학원, 정보통신공학과(공학석사)
 2013년 아주대학교 정보통신전문대학원, 정보통신공학과(공학박사)
 2020년~현재 CNAI 수석연구원
 관심분야 : 데이터베이스, etc.
 E-mail : keonsoo@cnaikr



김 중 연(Jung-Yeon Kim)

2009년 대구대학교 정보공학과(공학사)
 2013년 Lancaster Univ. Graduate School, Dept. of Mobile & Ubiquitous Computing (공학석사)
 2018년 순천향대학교 대학원 의료과학과(이학박사)
 2018년~현재 순천향대학교 ICT융합연구소 연구교수
 관심분야 : 지능형시스템, Mobile Health, 원격제할, etc.
 E-mail : betterwithme@sch.ac.kr