

소리데이터 수집 및 활용을 위한 개인정보보호 방안 연구

최지우, 이인우, 김은진

한국과학기술정보연구원

liewhite@kisti.re.kr, inulee@kisti.re.kr, kimej@kisti.re.kr

A Study on Privacy measures of sound data collection and uses

Choi Ji Woo, Lee In Woo, Kim Eun Jin

KISTI

요약

ICT의 비약적인 발전과 더불어 데이터의 중요성이 증대되고, 활용의 가치가 늘어나고 있다. 소리데이터를 수집하고 수집된 데이터들은 네트워크를 통해 저장되고 공유될 수 있다. 이로 인해 개인정보보호 측면에서 기존의 IT 환경 대비 새로운 문제들이 존재한다. 본 연구는 소리데이터 수집부터 활용하는데 발생 가능한 개인정보보호 관련 이슈를 파악하고 대응 방안을 법·제도적 및 기술적 방안을 제안하는 것을 목적으로 제안하였다.

I. 서론

ICT의 비약적인 발전과 더불어 데이터의 중요성이 증대되고, 활용의 가치가 늘어나고 있다. 그중 소리데이터는 다양한 응용 분야의 환경에서 활용하는 연구가 시도되고 있다. 소리데이터를 수집에 있어 가장 큰 문제점은 국내·외에서는 개인정보보호를 위해 개인정보 보호 법제를 갖추고 정보 주체의 동의 없는 개인정보의 수집과 활용을 엄격하게 금지하는 등 관리 및 보호 체계를 마련하고 있다.

그러나 인공지능 등 관련 기술이 발전하고 국가산업과 글로벌 경쟁력의 핵심이 됨에 따라 데이터 활용에 대한 수요와 관심도가 높아지는 상황이다.

데이터의 활용과 개인정보 보호의 균형을 유지하기 위한 방안으로 개인정보를 비식별화하여 사용하는 것에 대해서는 법률이 일부 규제적 조치를 완화하는 경향성을 보이고 있다. 이에, 국내에는 2020년 「개인정보보호법」을 개정하여 ‘가명 정보’라는 개념을 신설하고, 특정한 경우에는 정보 주체의 동의 없이 가명화된 개인정보를 사용할 수 있도록 허용한다.

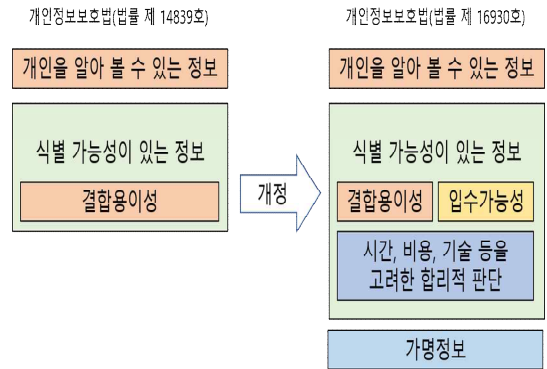
본 논문에서의 소리데이터란 ‘음파신호의 주파수 샘플을 추출하여 디지털로 변환된 소리 신호의 형태는 WAV, MP3, FLAC 등이 존재하며 활용 시 WAV 파일로 저장하여 보편적으로 활용 할 수 있게 전처리하며 수집된 데이터는 표준화 및 정의된 프로토콜을 통해 저장하고 공유 할 수 있다. 소리데이터 수집에 대해 일일이 개인에게 사전 동의를 받거나 특별한 상황 발생 시마다 개인들에게 매번 동의를 받는것이 여의치 않을 수 있다. 소리데이터 수집·활용시 개인정보가 포함되어 있을 경우의 대응 방안을 제시하고자 한다.

본 논문에서의 소리데이터란 ‘음파신호의 주파수 샘플을 추출하여 디지털로 변환된 소리 신호의 형태는 WAV, MP3, FLAC 등이 존재하며 활용 시 WAV 파일로 저장하여 보편적으로 활용 할 수 있게 전처리하며 수집된 데이터는 표준화 및 정의된 프로토콜을 통해 저장하고 공유 할 수 있다. 소리데이터 수집에 대해 일일이 개인에게 사전 동의를 받거나 특별한 상황 발생 시마다 개인들에게 매번 동의를 받는것이 여의치 않을 수 있다. 소리데이터 수집·활용시 개인정보가 포함되어 있을 경우의 대응 방안을 제시하고자 한다.

II. 본론

소리데이터의 수집에 있어 산업현장(기계음, 가스 누출 등) 비 언어에 대한 수집 및 활용에 있어 개인정보가 포함되지 않은 경우가 많으나, 백화점 및 대중교통 등의 공공장소에서 소리데이터 수집하기 위해서는 개인정보라고 할 수 있는 것은

개인과 개인과의 대화가 수집되면 목적에 맞지 않는 내용이 포함되기 때문에 개인정보에 대한 이슈가 발생하게 된다. 이를 해결하기 위해 수집 및 활용을 위한 방안을 제안하고자 한다. 개인정보보호법 제2조(정의)를 보면 가명처리란 개인정보의 가명처리는 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없는 특정 개인을 알아볼 수 없도록 처리하는 것을 말한다.



<fig 1. 개인정보보호법의 개정에 따른 개인정보 정의의 변화>

개인정보개념으로 소리데이터를 정의 하면 개인의 음성은 현행 「개인정보보호법」상 ‘개인정보’에 해당하므로 동의 없이 수집 불가하며 개인정보보호위원회는 ‘음성’을 ‘생체정보’로 보고 있다. 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보 없는 특정 개인을 알아 볼 수 없도록 처리하는 것을 ‘가명처리’라고 한다. 가명처리된 개인정보를 가명정보라고 한다. 이와 같이 소리데이터를 개인정보로 인식함으로 적법하게 수집한(정보 주체의 동의 등에 기하여) 개인정보를, 통계작성, 과학적 연구, 공익적 기록 보존 목적에 의해서만 가능하다. 즉, 소리데이터를 수집할 때 포함된 개인정보를 가명화하여 사용하기 위해서도 정보 주체의 동의나 법률상의 근거에 의하여 수집된 개인정보를 통계작성, 과학적 연구, 공익적 기록 보존의 목적이 있어야 한다.

그러나, 소리테이터를 수집하는 목적이 과학적 연구에 해당하므로 가명화의 목적 요건에는 부합할 수 있으나 당초에 정보 주체의 동의에 기하여 수집하지 않는 경우에는 가명화 자체를 진행할 수가 없다.

개인정보 수집·활용의 관련 문제와 대응 방안을 해결하기 위해 전제 조건으로 2가지로 ① 개인정보 수집·활용 동의 와 ② 소리테이터 내 개인의 음성이 ‘대화’에 해당하는 경우로 가정한다. ①에서 연구수행 목적의 개인정보 수집은 위의 정보 주체의 동의 외에 다른 규정에 부합해야 가능하다. 현재로서는 「과학기술기본법」 등에 관련 연구 수행을 위해 불가피하게 개인정보가 수집되는 것을 허용하는 근거를 마련하여야만 정보 주체의 동의 없이 수집·활용이 가능하다. 이 외에 개인정보를 동의 없이 수집·활용이 가능한 경우는 다음과 같다. 시간과 비용, 기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보일 경우 「개인정보보호법」 제58조의2에는 법 적용이 배제되므로 오랜 시간과 노력 등을 기울여야만 개인을 식별할 수 있는 수준의 데이터일 경우에는 동의 없이 수집·활용이 가능하다. 익명 정보 또는 「통계법」에 따라 수집되는 경우, 국가안전보장과 관련된 정보 분석을 목적으로 수집 또는 제공 요청되는 개인정보이거나, 공중위생 등 공공의 안전과 안녕을 위하여 긴급히 필요한 경우로 일시적으로 처리되는 개인정보, 언론·종교단체·정당이 각각 취재·보도·선교·선거 입후보자 추천 등 고유 목적을 달성하기 위하여 수집·이용하는 개인정보에 대해서는 동의 없이 수집·활용이 가능하다.

현재 소리테이터를 수집하는 목적은 연구 또는 공공데이터 수집 목적 등이라 해석 가능 하나 이에 의해서는 「개인정보 보호법」 상의 적용을 피하기 어려울 것으로 보인다. 수집하기 이전에 식별성이 완전히 제거되었거나 당초 오랜 시간과 노력 등을 기울여야만 개인이 식별 가능한 데이터라고 객관적으로 인정되는 경우에만 수집이 가능하다. 공개된 개인정보라 하더라도 공개한 정보 주체의 의도가 어느 누구에 의해서 활용될 수 있음을 인식하고 공개한 개인정보에 대해서만 수집·활용이 가능하고, 이 경우에도 개인정보 수집·활용의 정당성이나 개인정보 주체에 대한 이익이 있으나 여부, 공익적 필요성이 고도로 요구되는 개인정보에 해당하느냐 여부 등을 까다롭게 요구하고 있어 공개된 개인정보를 수집·활용하는 것도 어려울 것이다. ②개인과 개인의 대화와 통신은 「통신비밀보호법」에 따라 엄격히 보호되고 있다. 동 법 제3조 제1항은 “누구든지 이 법과 형사소송법 또는 군사법원법의 규정에 의하지 아니하고는 우편물의 검열·전기통신의 감청 또는 통신 사실확인자료의 제공을 하거나 공개되지 아니한 타인간의 대화를 녹음 또는 청취하지 못한다”라고 하고 있다. 개인 간의 대화는 절대적으로 보호되고 있는 것으로 형사사법상 증거 수집 등의 목적 이외에는 녹음을 할 수 없으며, 이를 위반한 경우 1년 이상 10년 이하의 징역과 5년 이하의 자격정지에 처하게 된다. 「통신비밀보호법」 제16조 따라서 소리테이터 수집 과정 중 ‘대화’가 포함되는 경우 「개인정보보호법」 외에 「통신비밀보호법」도 위반될 수 있음을 고려하여야 한다.

이러한 개인정보 관련 문제를 해결하기 위해서는 방안을 제시하고자 한다.

<fig2. 개인정보 관련 소리테이터 수집 및 활용 방안>

구분	소리테이터 수집 방안
1	개별 법률상 근거 마련
2	규제 샌드박스 이용
3	수집단계에서 음성정보 삭제 또는 노이즈 처리

1) 개별 법률상 근거 마련 관련 법률상에 ‘연구기관이 연구 목적의 소리테이터를 수집할 수 있는 법적 근거’를 두고 수집 단계에서 음성이 포함되더라도 수집이 가능하도록 규정한다. 법률의 체계상 연구 목적의 데이터 수집은 소리테이터에 한정하기보다 다양한 데이터를 포함하여 수집할 수 있는 근거를 두는 것이 적절해 보이며, 개인정보 수집이 목적 자체가 아닌 비 개인정보 수집 과정에서 부득이 수집되는 경우로만 한정한다. 다만 이 경우에도 개인정보 무단수집의 위험성에 대한 비판이 존재할 것이므로 수집 즉시 개인정보는 삭제하거나 노이즈 등 비식별 처리하도록 의무화하고 있다. 음성정보가 개인을 식별하는지를 제3의 독립기관이나 조직에서 관리·감독하도록 하여 활용은 하되 철저한 관리 기반으로 비판에 대응한다. 기타 개인정보 수집을 최소화하는 의무규정, 수집·파괴한 정보에 대한 기록관리 및 보고 의무화 등의 관리조치 등을 추가한다. 법률상 근거를 두는 것이 가장 적합하고 효과적일 것이나 실질적으로 실현될 가능성이 낮다는 단점이 존재한다. 특히 최근에는 공공장소 등에서 대량의 데이터를 빠르게 수집하고 처리하는 기술환경에 대한 우려가 높으며 특히 개인정보에 대해서는 거부감과 경계가 상당하다. 연구 목적으로 데이터를 수집하는 것에 대한 당위성과 이 과정에서 부득이하게 포함되는 개인정보만 수집, 처리한다는 점에 대한 설득이 가능하여야 할 것이다. 2) 규제 샌드박스는 일정 조건 아래 혁신적 신기술을 테스트할 수 있도록 규제를 유예 또는 면제해 주는 제도를 의미한다. 기간이 한정적이고 여러 조건 하에서만 실행이 가능하므로 장기적 연구 추진의 해결 방안이 될 수 없고 장애 요인이 많을 것으로 예상된다. 3) 데이터를 수집하는 단계에서 개인정보의 식별성을 제거하는 방법을 고려해볼 수 있다. 다만 기술적으로 구현하는 것에 상당한 시일이 소요될 것으로 예상되며, 식별성을 제거하였다 하더라도 오류 등이 발생하거나 완전히 식별성을 제거하지 못한 경우에는 다시 법률 위반의 결과에 도달하므로 완전히 기술력을 확보하지 못하지 않는 이상 위험성이 존재한다. 이와 같이 소리테이터 수집 및 활용에 개인정보와 관련된 문제를 해소 방안을 제시하였다.

III. 결론 및 제언

본 논문에서는 소리테이터의 수집 및 활용을 위해서는 개인정보개인정보 관련 문제를 인식하고 활용하는데 필요한 법·제도적인 근거 및 기술적인 방안을 제시하였다. 또한 소리테이터를 처리 하기 위해서는 설계단계부터 프라이버시를 고려하는 개념을 제도적으로 적용하는 방안은 국외 법/제도 제정 현황을 참조하되 국내 실정을 맞게 도입 검토가 필요하며, 윤리적 기준을 철저히 준수해야 한다. 사용자 데이터의 민감성을 인식하고, 이를 보호하기 위한 강력한 정책과 절차가 수반되어야 한다. 이러한 절차를 바탕으로 연구 과정에서 생성된 데이터 및 도구를 오픈 소스로 공유함으로써, 지식의 공유를 촉진하고, 연구 커뮤니티 전반의 협력을 강화할 수 있다.

ACKNOWLEDGMENT

이 논문은 2023년도 한국과학기술정보연구원(KISTI)의 자체사업으로 수행된 연구입니다.(과제번호: J-23-NB-C01-S01)

AI, Sound Intelligent, Data, Privacy

참 고 문 헌

- [1] 개인정보보호법
- [2] 사물인터넷(IoT)환경에서 개인정보보호 강화를 위한 제도 개선 방안 (정보보호학회 논문지, 2016.08), 이애리, 손수민, 김현진, 김범수