

## 한국 수사기관의 범인음성식별절차 도입을 위한 제언

배 민 선

한림대학교 심리학과

이 미 선<sup>†</sup>

한림대학교 범심리학연구소

범인음성식별(voice identification)은 범인과 용의자의 목소리를 대조하여 범인을 식별하는 과정을 의미한다. 지금까지 음성식별과 관련된 연구 결과에 따르면 음성식별과정은 오류를 포함할 수 있으며 특정 조건 하에서 정확성이 감소될 수 있음을 확인하였다. 예를 들어, 범인음성식별과정에서 제시되는 적성수준의 음성 샘플의 수와 제시된 목소리의 길이는 정확성을 향상시킬 수 있는 반면, 목격 시점과 라인업 사이의 시간적 간격이 길어질수록 정확성은 감소한다. 목격자의 인종이나 연령, 성별 및 확신도 역시 음성식별의 정확성에 영향을 줄 수 있다. 영국의 경우, 일찍 범인의 목소리에 대한 진술의 중요성과 범인음성식별절차에 대한 공식지침의 필요성을 깨닫고 자세한 공식지침을 권고하고 있다. 하지만 우리나라의 경우, 범인음성식별과 관련된 인식은 여전히 미미한 수준이다. 범인음성식별의 신빙성 확보를 위한 표준화된 범인음성식별절차의 도입을 위해서는 지속적인 연구와 더불어 관련 실무자의 관심이 요구된다.

주요어 : 범인음성식별절차, 음성목격자, 목격자 식별, 음성식별, 목격자 진술

---

<sup>†</sup> 교신저자 : 이미선, 한림대학교 범심리학연구소, 강원도 춘천시 한림대학길1  
E-mail : misunyi23@gmail.com

범인음성식별(voice identification)이란 사건 당시 범인의 목소리를 들은 목격자 또는 피해자(이하 목격자)가 자신의 기억 속 범인의 목소리와 용의자의 목소리를 대조하여 범인을 식별하는 하는 절차를 의미한다. 두 개의 음성 샘플의 동일성을 판단하는 성분분석과는 다르게 음성식별은 목격자 기억의 신빙성과 관련되어 있다는 점에서 오히려 용의자의 인상착의 등에 의한 목격자의 범인식별절차(eyewitness identification procedure)와 유사성을 갖는다. 음성식별과정은 과거 경험한 사건의 기억을 기초로 하기 때문에 사람의 인지능력과 지각능력에 의해 영향을 받을 수 있으며 목격한 당시 또는 진술한 당시의 상황에 따라 상당한 오류를 포함할 수 있기 때문에 식별절차에 영향을 미칠 수 있는 요인에 대해 고려해야 할 것이다(Bull & Clifford, 1984). 영국 등의 나라에서는 일찍이 범인식별절차의 위험성을 고려하여 목격자 진술의 오염 가능성을 최소화하고 정확성을 향상시킬 수 있는 표준화된 지침을 개발하여 실무에서 활용하고 있다(Wilcock, Bull, & Milne, 2008). 그러나 우리나라의 경우, 범인음성식별절차에 대한 공식적인 지침이 존재하지 않으며, 활용 실태에 대해서도 확인된 바 없다. 사실상 범인음성식별절차는 우리나라에서 여전히 생소한 개념으로 관련 연구조차 거의 존재하지 않는다. 이에 본 연구의 목적은 다음과 같다. 첫째, 우리나라에서 범인음성식별이 쟁점이 되었던 판례를 통해 범인음성식별절차 도입의 필요성 및 범인음성식별 개념을 소개한다. 둘째, 범인음성식별절차 관련 연구를 개관하여 식별절차에 영향을 미치는 요인에 대해 확인한다. 셋째, 영국의 범인음성식별절차를 소개하고 마지막으로 우리나라에서 범인음성식별절차의 도입 시 고려해야

할 사항에 대해 논의하고자 한다.

### 범인음성식별(voice identification)의 소개

일반적으로 범죄 사건에서는 범죄 현장에서 찾은 피의자의 지문, 모발, 정액, 범행 도구 등 객관적인 물적 증거가 존재하며, 이러한 경우에 범인식별은 피의자의 진술에만 의존할 필요가 없다. 다만 객관적인 증거가 존재하지 않거나 존재하더라도 증명력이 확보되지 않는 경우, 목격자의 진술은 사건 해결에 있어서 유일한 증거가 되며, 이때 진술의 신빙성은 유무죄 판단에 결정적인 역할을 하기도 한다. 더욱이 객관적 증거의 부재와 동시에 사건 당시, 현장이 어두웠거나 얼굴을 가려 제대로 범인의 얼굴을 보지 못한 경우, 범인의 말씨, 목소리, 억양 등은 범인을 식별하는데 결정적인 요인으로 작용할 수 있다. 현재 우리나라에서 범인음성과 관련된 목격 진술이 쟁점이 되었던 사건에 대한 법원의 증거 능력 판단은 다음과 같다.

대법원 1984.6.26. 선고 84도947 판결

본 사건은 공갈미수사건으로 피해자는 협박범이 경상도 사투리를 사용하는 25~26세 정도의 청년인 것 같았다고 진술하였으며 피고인이 경찰과 검찰에서 한 진술을 토대로 음성의 굴곡, 억양을 봤을 때 범인의 목소리와 동일한 목소리라고 증언하였다. 하지만 대법원은 경상도 사투리가 목소리 억양의 굴곡이 있는 것이 특징이지만 동일인이라고 단정할 증거로는 부족하며 협박편지 등의 필적증거도 동일인이 작성했다고 단정 지을 수 없다고 판

단하고 원심의 유죄 판결을 파기하였다.

서울고등법원 1998.7.8. 선고 97노1260  
판결

범인이 피해자의 집에 침입하여 강간을 한 사건이다. 범인이 피해자의 집에 침입할 당시 전등을 모두 꺼 방이 어두웠고 범인이 수건으로 얼굴을 가리고 있어 피해자가 제대로 얼굴을 보지 못했다. 이후 피해자는 피고인의 목소리와 억양이 범인의 것과 같다는 취지로 진술하였으나 고등법원은 피해자가 범인의 얼굴을 제대로 보지 못했고 피해자가 다른 암시로부터 영향 받기 쉬우며 범인을 관찰한 객관적인 조건도 좋지 않았다는 점에 주목하였다. 또한 어둠 속에서 부엌칼로 위협당하며 강간을 당한 피해자가 범행과정 중 범인과 대화하며 들은 목소리로 범인의 목소리를 정확하게 기억하고 있다는 것도 수궁하기 어렵다는 점을 들어 피해자의 범인식별진술과 정황 증거만으로는 피고인이 범인이 아닐지 모른다는 의구심을 떨쳐버릴 수 없으며 혐의를 인정할 만한 증거가 없다는 이유로 피고인에 대한 무죄 선고가 정당하다고 보고 항소를 기각하였다.

대전고등법원 2008.6.18. 선고 2008노50  
판결

범인은 피해자의 집에 침입해 안방에 잠들어 있던 피해자에게 다가가 과도로 위협하며 조용히 하지 않으면 죽이겠다며 협박하고 피해자의 얼굴을 이불로 덮는 등의 방법으로 피해자의 반항을 억압하고 지갑을 강취한 후 피해자를 강간하려고 했으나 미수에 그쳤다. 피

해자는 사건 당시 실내등이 꺼져 있어 범인의 얼굴을 분명하게 목격하지는 못하였으나 범인의 목소리는 맑고 매끄러운 편이었다고 진술하였다. 피해자는 피고인과 대면 후, 다른 인상착의는 범인으로 보기 어렵지만 목소리가 범인과 놀랄 정도로 같다고 진술하였다. 원심에서는 사건현장에서 발견된 담배꽂초의 유전자와 피고인의 유전자가 일치한다는 유전자 검사결과에 의해 피해자가 피고인이 범인일 것이라는 암시를 받아 피해자의 기억이 왜곡되었을 가능성이 있다는 점을 문제 삼았다. 또한 범인식별절차 이전에 수사기관에서 범인의 인상착의에 관한 피해자의 진술을 상세히 기록하지 않았으며 식별과정에서도 피고인과 인상착의가 유사한 다른 사람들을 동시에 대면하거나 그들의 사진을 동시에 제시받은 상태에서 범인을 지목하게 한 것으로 보이지 않는다고 판단하였다. 또한 그 과정과 결과를 서면화한 자료도 찾아볼 수 없어 이 사건에서 범인식별절차에서의 준수사항을 제대로 지켰다고 하기 어렵다고 보았다. 피해자 진술의 신빙성을 인정한다고 하더라도 체격, 목소리 등만을 범인의 유력한 특징으로 삼는다면 용의자의 범위가 너무 넓어져 범인을 피고인으로 단정할 수 없다는 입장을 보였다. 그리고 증거로 수집된 담배꽂초에 대해서도 수집과정에서 사진, 문서 등 다른 기록이 없었고 담배꽂초의 출처가 확실치 않다고 보아 유전자 감정결과만으로 피고인이 본 범행을 저지른 범인이라는 것이 합리적 의심의 여지가 없을 정도로 충분히 증명되었다고 보기 어렵다고 판단하여 일부 무죄를 선고하였다. 반면 고등법원은 반복된 피해자의 진술 내용이 구체적이고 일관되어 진술의 신빙성을 충분히 인정할 수 있다고 보았다. 또한 범인이 피해자를 협

박하는 과정에서 범인의 목소리 특징을 충분히 파악할 수 있었던 것으로 볼 수 있는 점에 비추어 보면 피고인과 인상착의가 유사한 다른 사람들을 동시에 대면한 상태에서 범인을 지목하도록 하는 등의 적법한 범인식별절차가 이루어지지 않은 것만으로는 피해자의 진술에 신빙성이 없다고 단정하기 어렵다고 보았다. 또한 비록 공판에서 대면한 피고인에 대해 광대뼈와 팔뚝의 굵기 등이 범인과 다르다고 진술하였으나 사건 당시 공포와 긴장 상태에 있던 피해자가 범인의 손목 등의 두께를 정확히 파악하여 기억하는 것은 자연스럽다고 할 수 없고 피고인의 체중증감 등을 고려하면 피고인의 대체적인 인상착의와 목소리가 범인과 비슷하다는 피해자 진술의 신빙성을 탄핵하기에 부족하다고 보았다. 또한 사건 현장에서 발견된 담배꽂이의 유전자 감정결과, 피고인과 유전자가 일치하였기 때문에 고등법원은 원심판결을 파기하고 징역 7년을 구형하였다.

#### 피해자의 범인식별진술에 대한 법원의 판단

앞서 살펴본 사건은 범인의 목소리를 전화로 들은 공갈 미수사건으로 범인의 얼굴을 보지 못한 상황이었으며 나머지 두 사건은 강간 사건으로 당시 사건 현장이 어두웠거나 얼굴을 가려 제대로 범인의 얼굴을 보지 못한 경우였다. 목격자의 범인과 피고인에 대한 인상착의 혹은 음성에 대한 동일성이 쟁점이 되었으며 목격 진술의 신빙성에 대한 판단이 사건의 유무죄 판결에 있어서 결정적인 역할을 한 것으로 보인다. 다만 목격자의 기억에 기초한 음성식별진술의 신빙성에 대한 법원의 판단은 부정적인 것으로 보인다. 특히 처음 소개된 두 사례(대법원 1984.6.26. 선고 84도947 판

결; 서울고등법원 1998.7.8. 선고 97노1260 판결)에서 비록 피해자는 범인과 용의자의 목소리가 유사하다는 취지로 진술하였으나 피해자 진술의 왜곡 가능성을 배척하기 어려우며 합리적 의심이 없을 정도로 충분히 증명되었다고 보기는 어려운 점 등의 이유로 피고인에게 무죄가 선고되었다. 세 번째 판례의 경우에도 하급심에서는 피해자의 암시여부와 수사기관에서 범인식별절차를 준수하지 않았음을 지적하며 피해자의 진술의 신빙성을 인정하더라도 그 진술만으로 피고인이 범인이라는 점이 충분히 증명되지 않는다고 보았다. 다만, 고등법원의 판결에서는 적법한 범인식별절차가 제대로 이루어지지 않은 것만으로 피해자의 진술을 탄핵하기에는 부족하며 피해자의 진술 내용이 구체적이고 일관되어 진술의 신빙성을 인정하였다. 위의 판례들은 음성식별과정에서 인간의 기억의 한계로 인한 오류 가능성을 인정하고 있으며 수사기관에서 적법한 범인식별절차가 이루어지지 않았다면 해당 진술의 증명력을 확보할 수 없음을 명시하고 있다. 따라서 피해자의 범인음성식별진술을 확보하는 과정에서 적합한 절차를 준수하지 않는다면 증거효력을 잃게 될 가능성이 있다. 하지만 현재 우리나라에는 음성에 대한 목격진술의 신빙성 판단 기준이 모호하며 음성을 통한 범인식별과정에서 목격진술의 증명력을 확보할 수 있는 객관적이고 표준화된 식별절차에 대한 합의는 존재하지 않는다.

#### 범인음성식별의 개념

‘범인음성식별’이란 범인의 목소리를 들은 목격자가 용의자와 들러리(범인식별절차에서 용의자 외에 제시되는 사람)들의 목소리를 들

고 기억 속에 있는 범인의 목소리와 대조하여 그와 같은 목소리인 사람을 찾는 절차이다(Hollien, 2012). 범인음성식별은 범인과 용의자 간의 음성을 비교하여 동일성을 확인한다는 점에서 성문분석과 유사하다고 볼 수 있다. 성문분석은 성문(聲紋, voice print)을 이용하는 방법으로 녹음된 목소리의 성문과 비교하고자 하는 사람의 목소리의 성문을 비교·대조하여 동일성 여부를 판단하는 것이다. Kersta(1962)에 따르면 사람의 목소리에는 지문처럼 성문이라는 독특한 특징이 존재한다. 이 성문은 음향분석기를 통해 목소리의 음파 등을 분석하여 그 결과를 시각화한 것이다. 이것은 개인마다 특징을 지니고 있어 용의자와 범인의 동일성 여부를 가리는 중요한 단서가 되고 있다. 특히 사람의 음성은 각 개인의 서로 다른 음성기관 즉, 성대, 치아, 비강 등의 크기, 모양의 물리적인 특성에 따라 결정되며 언어 습득 시의 사회·언어적 환경과 개인의 말하는 습관 등에 의해 발음, 어휘, 억양 등이 영향을 받아 다르게 형성되기 때문에 성문의 분석결과를 종합하면 동일인인지 판단할 수 있으며 또한 그것을 토대로 성별, 연령, 출신지 등을 추정할 수 있다(홍수기, 2006).

반면, 음성식별절차는 범인의 목소리가 목격자의 기억에만 존재하여 용의자의 목소리를 듣고 피해자가 자신의 기억 속 범인의 목소리와 비교·대조하여 같은 목소리를 찾아내는 것으로 범인의 얼굴을 목격한 목격자가 사진 혹은 실물을 보고 동일인 여부를 판단하는 범인식별절차(eyewitness identification)와 유사한 과정을 따른다. 범인식별절차란 범인의 얼굴을 목격한 목격자 또는 피해자가 용의자를 직접 보거나 사진, 비디오 등을 통해 본 다음 목격한 범인의 얼굴과 일치하는지 식별하는 과정이다.

범인식별은 기억과 관련되어 있기 때문에 인간의 기억 과정, 개인의 지각능력, 목격 당시의 상황, 범인식별절차 등에 의해 영향을 받을 수 있으며, 식별의 오류 가능성을 내재하고 있다(McGrath & Turvey, 2014). 범인음성식별절차 역시 시각적 범인식별절차와 유사하게 인간의 기억에 의지하고 있기 때문에 식별 정확성은 다양한 요인에 의해서 영향을 받을 수 있다. 따라서 목격자가 기억을 회상하는 것에 도움을 줄 수 있는 표준화된 절차를 만들어서 실시하는 것은 상당히 중요한 부분이라고 할 수 있다.

### 범인음성식별의 신빙성에 영향을 미치는 요인

객관적 증거가 확보되지 않은 상태에서 목격자 진술은 사건의 유무죄 판단 및 피고인의 범인 동일성 식별에 있어 결정적인 역할을 한다. 일반적으로 목격자의 범인식별 증거는 유무죄 판단에 있어서 강력한 증거로 고려되며(Kassin & Neuman, 1997), 특히 목격자가 높은 확신감을 가지고 범인식별을 한 경우, 유죄 선고가 되는 가능성이 높아진다(Wilcock et al., 2008). 하지만 목격진술은 사실 확인에 있어서 결정적인 역할을 하는 동시에 오판을 이끌어 낼 위험성 역시 매우 높다. 실제로 영국과 미국에서 실시된 연구 결과에 따르면 부정확한 목격자 식별이 오판의 주요 원인인 것으로 나타났다(Bedau & Radelet, 1987; Huff, Rattner, & Sagarin, 1986; Rattner, 1988). 목격자 진술은 사건에 대한 개인의 기억에 기초하기 때문에 실제 진실과는 차이가 있을 수 있으며, 외부적 요인에 의한 오염 가능성이 존재한다. 특

히 정보가 입력, 저장, 인출되는 인간의 기억 과정에서 다양한 요인이 정확성에 영향을 줄 수 있는데 개인의 사건에 대한 기억은 객관적 실체이기 보다는 개인이 세상에 대해 갖고 있는 일반적인 정보를 사용하여 재구성되기 때문에 개인의 경험이나 고정관념 등에 의해 왜곡될 수 있다(Milne & Bull, 1999). 또한 목격시기와 면담 사이의 지연이 길어질수록 세부사항은 소실될 수 있으며, 외부 원인에 의해 오염될 여지가 커져 그 정확도는 감소할 수 있다(Meissner & Brigham, 2001). 따라서 신뢰성 있는 범인식별을 위해서는 목격자의 능력을 극대화하는 반면, 진술의 오염을 초래할 수 있는 외적 요인을 최대한 차단한 표준화된 절차에 의해 범인식별절차를 실시해야 할 것이다.

Wells(1978)는 범인식별의 신빙성에 영향을 미칠 수 있는 변인들을 형사사법체계에 의한 통제 가능성 여부에 따라 평가자 변인(estimator variables)과 시스템 변인(system variables)으로 분류하였다. 평가자 변인은 목격자 또는 가해자의 특성이나 상황적 요인에 따라 정확성에 영향을 미칠 수 있는 요인으로 형사사법체계에 의해 통제되지 않는다(Narby, Culter, & Penrod, 1996). 예를 들어, 목격자의 연령이 너무 어리거나(Pozzulo & Lindsay, 1998) 60세 이상인 경우(Searcy, Barlett, & Memon, 1999), 젊은 성인에 비해 범인식별능력이 떨어지는 것으로 나타났으며, 가해자가 여성인 경우, 남성에 비해 더 쉽게 식별되는 경향이 있다(Shapiro & Penrod, 1986). 그리고 일반적으로 자신과 동일한 성별에 대해서는 더 잘 식별하였다(Shapiro & Penrod, 1986). 목격자가 식별과정에서 특정 용의자를 보고 범인이라고 매우 확신하는 상황에서도 해당 진술은 부정확할

수 있다. 이는 사람들이 자신의 기억을 모니터링 하는 것에 실패하거나 실제 정확도와는 무관한 근거 없는 확신 경향 가능성 때문이다(Busey, Tunnicliff, Loftus, & Loftus, 2000). 이외에도 목격 당시 스트레스 수준이 높아지면 식별의 정확성이 떨어질 수 있으며(Morgan et al., 2004; Valentine & Mesout, 2009), 가해자가 총이나 칼 등 위협적인 물건을 들고 있는 경우, 목격자의 주의가 흉기에 집중되어 오히려 범인식별의 정확성이 감소되기도 한다(Loftus, Loftus, & Messo, 1987).

평가자 변인과는 달리 형사사법체계의 통제 하에 있는 시스템 변인은 라인업(line-up)과 관련되어 있다. 라인업(line-up)이란 용의자와 다른 사람들(foils)을 목격자에게 제시한 후 여러 사람 중에서 범인을 지목하도록 하는 일련의 과정을 의미한다. 기존 연구에서는 라인업에서 제시되는 사람의 수와 제시방법, 제시할 때 지시사항 등에 따라서 범인식별의 정확성에 영향을 미칠 수 있음을 확인하였다. 라인업 크기와 관련해서 라인업에 제시되는 사람의 수가 증가할수록 용의자가 우연에 의해 지목될 가능성이 낮아지기 때문에 반드시 적정수 이상의 들러리(foils)가 필요하다(Wilcock et al., 2008). 최소한의 들러리 수는 3~5명이 필요하며(Nosworthy & Lindsay, 1990), 일반적으로 용의자 이외에 8명 정도가 적정한 것으로 고려된다(Advice on the use of voice identification parades, 2003) 이때 들러리들은 용의자와 연령이나 신장, 외모 등이 유사해야 한다(Luus & Wells, 1991). 용의자 이외에 여러 명의 들러리가 존재하는 경우에도 이들의 인상착의 및 특징이 용의자와 매우 차이가 있어 라인업 상에서 용의자가 두드러지게 된다면 기능적으로 들러리는 부재한다고 볼 수 있다(예를 들어,

용의자만 서양인이고 다른 들러리들은 동양인인 경우).

제시 방법 역시 범인식별의 정확성에 영향을 미치는 것으로 나타났다. 식별절차의 제시 방식은 동시적 제시와 순차적 제시가 있다. 동시적 제시 방법은 라인업 상의 모든 용의자 및 들러리들을 동시에 제시하는 것이고 순차적 제시는 한 명씩 순차적으로 제시하는 것을 의미한다. 동시적 제시는 라인업의 모든 인원을 동시에 보여주는 방식으로 라인업의 모든 사람들을 상대적으로 비교하며 목격자의 기억 속 범인과 가장 닮은 사람을 선택하는 상대적인 판단 전략을 사용하게 되는 제시 방식이다. 반면, 순차적 제시 방식은 라인업의 인원들을 한꺼번에 보는 것이 아닌, 한 번에 한 명을 보며 목격자의 기억 속 범인과 대조하는 절대적인 판단 전략을 사용하게 하는 방식이다. 여러 연구에서 동시적 제시에 비해 순차적 제시가 절대적인 판단 전략을 통해 오식별을 줄일 수 있는 것으로 나타났다(Kneller, Memon, & Stevenage, 2001; Lindsay & Wells, 1985).

범인식별절차를 실시할 때 목격자에게 라인업에 가해자가 있을 수도 있고 없을 수도 있다는 비편향 지시는 범인식별진술의 정확성에 영향을 미치는 것으로 나타났다. Malpass와 Devine(1981)에 의하면 목격자에게 편향 지시를 했을 경우, 범인이 없는 라인업에서 오식별을 할 확률이 78%였지만 비편향 지시를 한 경우에는 33%로 떨어졌다. 라인업 지시사항에 관한 18개 연구에 대해 메타 분석한 Steblay(1997)의 연구결과에서도 이와 유사하게 범인이 없는 라인업에서 비편향 지시를 하지 않은 경우(편향 지시), 정확성이 떨어지는 것으로 나타났다. 특히 라인업의 용의자가 무고하다면 비편향 지시는 매우 중요하게 작용할

수 있음을 보여준다. 지금까지 대부분의 범인식별 관련 연구는 시각적 식별절차에 초점을 두고 있으며 음성식별 관련 연구는 상대적으로 많이 이루어지지 않았다. 비록 음성식별절차의 많은 부분은 시각적 식별절차와 유사한 특성을 가지고 있으나 음성식별절차는 얼굴이 아닌 음성에 대한 식별이라는 점에서 차이가 있으며 이에 음성식별에 영향을 미칠 수 있는 변인을 확인하는 것은 중요하다.

#### 범인음성식별절차에서 사용되는 목소리의 개수

범인음성식별절차에서 제시되는 목소리의 적정 개수에 대해 연구자들마다 약간의 차이를 보이나 대체적으로 범인음성식별절차에서 총 6개 이상의 음성을 사용하는 것을 권고하고 있다. Broeders와 Amelsoort(1999)에 따르면 들러리들의 목소리는 최소 5개를 주장하였으며 Hollien(2012)은 음성식별절차에서 필요한 목소리의 개수는 용의자를 포함하여 6~8개를 사용할 것을 권고하고 있다. Yarmey(2007) 역시 하나의 용의자 목소리와 5명의 들러리 라인업을 구성하는 것을 권장하고 있다. 영국의 범인음성식별절차의 공식지침인 'Advice on the use of voice identification parades'(2003)에 따르면 라인업에는 총 9명의 발화자를 포함할 것을 권고하고 있다.

#### 제시 방식

음성식별절차에서는 목격자에게 여러 음성을 동시에 제시하여 인식하게 하는 것이 불가능하다. 따라서 순차적인 제시를 통한 식별절차를 진행하는 것이 필요하다.

### 제시되는 목소리의 길이(Duration)

제시되는 목소리의 길이 역시 중요하다. 식별해야 하는 음성의 길이가 지나치게 짧은 경우, 식별이 불가능할 수 있으므로 일정 길이 이상의 음성 샘플이 필요하다. 기존 연구 결과, 제시한 목소리의 지속시간이 70초인 조건이 30초인 조건에 비해 범인이 없는 라인업에서 범인이 존재하지 않음을 더 정확하게 식별해낸 것으로 나타났다(Kerstholt, Jansen, Van Amelsvoort, & Broeders, 2004). 또한 Yarmey와 Matthys(1992)의 연구 결과, 용의자가 있는 라인업에서 짧은 길이의 목소리 샘플(18초, 36초)에 비해 긴 길이의 목소리 샘플(120초, 6분)에서 더 우수한 정식별(hits) 수행을 나타냈다. 그리고 영국 범인음성식별절차의 공식지침에서는 샘플의 길이를 각 음성마다 60초로 정하고 있다(Advice on the use of voice identification parades, 2003).

### 파지 간격(Retention intervals)

목격 시점과 라인업 실시까지의 시간적 간격은 정확성에 영향을 줄 수 있다. 런던 경찰청에서 실시된 314건의 라인업 분석 결과, 사건 후 범인식별절차까지 6개월 이상의 지연이 있었던 경우, 정식별률은 34~46%로, 일주일 안에 이루어진 범인식별절차의 정식별률(65%)에 비해 감소하는 것으로 나타났다(Valentine, Pickering, & Darling, 2003). 음성식별절차의 경우, Clifford(1980)는 44명의 실험참가자를 대상으로 음성식별의 파지간격에 대한 연구를 실시하였다. 연구 결과, 파지 간격이 10분인 경우, 정식별은 40.8%이었으며, 24시간 20%, 7일 23.25%, 그리고 14일 19.25%로 파지 간격이

증가할수록 정식별은 감소하였다. 또 다른 연구에서는 오식별률(false alarm rate)에 대해서도 확인하였다. 범인이 존재하는 조건에서 목소리 샘플의 길이와 파지간격의 상호작용이 나타났다. 36초 길이의 목소리 샘플과 6분 길이의 목소리 샘플에서 오식별률이 유의미하게 달랐다. 36초 목소리 샘플의 경우, 즉시 테스트한 경우보다 일주일 뒤에 테스트한 경우에 더 많은 오식별이 있었으며 6분 목소리 샘플의 경우, 즉시 테스트한 경우보다 24시간 지연 테스트에서 오식별이 더 많았으나 일주일 지연 테스트의 경우에는 차이가 없었다(Yarmey & Matthys, 1992). 종합해보면 대체적으로 시간이 지날수록 식별의 정확도가 떨어짐을 알 수 있다. 따라서 가해자의 목소리/발화 프로파일을 구성하기 위해 사건 발생 후, 목격자를 가능한 일찍 인터뷰해야 하며 용의자가 체포된 후 가능한 빨리 음성식별절차를 실시해야 할 것이다(Yarmey, 2007).

### 비편향 지시(Unbiased line-up instructions)

비록 범인음성식별절차에서의 비편향 지시에 대한 연구는 존재하지 않았으나 앞서 살펴본 바와 같이 비편향 지시는 음성식별절차에 있어서도 매우 중요할 것으로 예상된다. 따라서 범인의 음성이 존재하지 않을 수 있다는 비편향 지시는 범인음성식별절차를 진행할 때도 고려해야 할 것이다. 영국 범인음성식별절차의 공식지침에서 식별 담당 경찰관은 목격자에게 재생되는 목소리 샘플 중에 용의자의 목소리가 있을 수도 있고 없을 수도 있다는 것을 설명해주어야 한다고 명시하고 있다.



### 변조(Disguise)

머리 스타일, 수염, 안경의 변화와 같은 얼굴에서의 큰 변화는 재인 수행을 거의 우연한 수준으로 감소시키는데 영향을 미치는 것으로 나타났다(Patterson & Baddeley, 1977). Cutler, Penrod와 Martens(1987)는 범인의 위장과 관련된 연구에서 가해자가 니트 풀오버 모자를 쓰거나 쓰지 않은 경우를 비교하였는데 범인이 모자를 쓰지 않은 경우에 식별 정확성이 더 높았음을 밝혔다. Pozzulo와 Marciniak(2006)의 연구에서는 동시적, 순차적 라인업 모두에서 머리 스타일의 변화는 정식별률을 떨어뜨렸다고 보고하였다.

Hollien, Majewski와 Hollien(1974)은 범인의 목소리 변조에 대한 연구를 실시하였다. 실험참가자가 발화자를 아는 경우에는 모든 말하기 조건에서 높은 정확도를 나타냈으며 음성 변조된 발화를 들은 경우에도 79%의 정확률을 나타냈다. 다만, 정상적인 발화에 대한 정확도가 거의 100%에 가까웠던 것에 비하면 음성 변조는 범인음성식별에 영향을 미치는 것으로 보인다. Procter와 Yarmey(2003)의 연구에 의하면 실험참가자들은 보통의 톤인 목소리와 속삭인 목소리 둘 중에 하나를 들었을 때 속삭인 목소리보다 보통의 톤인 목소리를 유의미하게 더 잘 식별하였으며 보통의 톤인 목소리에 대해 정확도와 확신도 사이의 유의미한 정적 관계가 발견되었다. Orchard와 Yarmey(1995)는 처음 범인의 목소리를 보통의 톤으로 들은 실험참가자들에게는 범인음성식별절차에서도 보통의 톤으로 목소리를 들려주었으며 처음 범인의 속삭이는 목소리를 들은 참가자들에게는 속삭이는 목소리를 제시하는 식별절차와 보통 톤의 목소리를 제시하는 식별절차, 둘

중에 하나를 실시하였다. 그 결과, 속삭이는 것을 통한 목소리 변조는 용의자가 있는 조건과 없는 조건의 라인업 모두에서 식별수행에 유의미한 영향을 미쳤다. 특히, 범인의 속삭이는 목소리를 최초에 들은 참가자에게 보통의 톤인 목소리로 라인업을 실시한 경우, 식별수행이 좋지 않았다. 따라서 범죄현장에서 본 범인의 모습과 들은 음성이 변장한 것이거나 변조된 것이었다면 그 이후의 식별수행에 부정적인 영향을 미칠 수 있다.

### 인종(Race)

음성식별과 인종에 관한 Vanags, Carroll과 Perfect(2005)의 연구에 따르면 실험참가자들은 자신과 같은 인종의 목소리를 식별하는 것(정식별률 62%)을 다른 인종의 목소리를 식별하는 것(정식별률 26%)보다 유의미하게 더 잘 수행하는 것으로 나타났다. 일반적으로 시각적 식별에 있어서도 용의자와 같은 인종의 목격자에 의해 식별되는 경우가 서로 다른 인종에 의해 식별되는 경우보다 더 정확한 것으로 나타났다(Valentine et al., 2003).

### 연령(Age)

Pozzulo와 Lindsay(1998)는 목격자의 연령과 목격자 진술의 정확성에 대한 다양한 연구들의 메타분석을 실시하였다. 연구 결과, 성인은 취학 전 아동(평균 연령 4세)과 비교하여 더 정확한 식별을 하는 경향이 있으나 5세 이상의 아동들은 성인들과 비교하여 정식별률에서 유의미한 차이를 나타내지 않았다. 또한 성인들은 모든 연령의 아이들보다 식별대상(target)이 존재하지 않는 라인업에서 정기각을 더 잘

하였다고 보고하고 있다. Valentine 등(2003)의 연구에 따르면 목격자의 나이는 용의자의 식별에 유의미한 영향을 미치는 것으로 나타났다. 30세 이하일 경우, 용의자를 더 잘 식별하는 경향이 있었으며 20세 이하 목격자의 48%가 용의자를 식별한 것에 비해 40세 이상 목격자의 28%만이 용의자를 식별한 것으로 나타났다.

목격자의 연령과 관련된 음성식별 정확성에 대한 연구도 다양한 결과를 나타내고 있다. Mann, Diamond와 Carey(1979)에 따르면 익숙하지 않은 목소리 재인에 대한 수행을 비교하였을 때 나이에 따라 두드러진 차이를 보였다. 특히 6~10세 사이에 정식별률이 급격히 증가했는데 6세의 수행은 우연의 수준에 머물렀으며 10세의 수행은 어른의 수행 수준에 근접했다. Öhman, Eriksson와 Granhag(2010)은 7~9세, 11~13세, 성인의 세 집단으로 나누어 익숙하지 않은 목소리에 대한 식별 정확성의 연령별 차이를 알아보았다. 모든 실험참가자들의 반은 식별대상(target)이 있는 조건, 나머지 반은 식별대상(target)이 없는 조건에 할당되었다. 참가자들은 모든 조건에서 좋은 수행을 보이지는 못했으나 세부적으로 보면 모든 연령 그룹은 식별대상(target)이 없는 조건에서 우연한 수준 이상으로 수행을 나타냈다. 반면, 식별대상(target)이 있는 조건에서는 11~13세 그룹이 유의미하게 우연의 수준 이상인 27%의 정식별률을 나타냈고 나머지 그룹은 우연한 수준보다 유의미하게 높은 수준의 수행을 나타내지는 못했다.

#### 성별(Gender)

Valentine 등(2003)의 연구에서 목격자의 성별

은 라인업에서 용의자가 아닌 들러리를 선택하는 오식별에만 유의미한 영향을 미치는 것으로 나타났다. 즉, 용의자 식별에 있어서는 남녀의 정식별률이 거의 동일하였지만 오식별률에 있어서는 남성이 여성보다 오식별을 할 확률이 적었다는 것이다. 반면, 음성식별과 관련된 연구에서 남성 실험참가자들은 여성의 목소리와 남성의 목소리를 재인하는 것에 있어서 유의미한 차이를 보이지 않았으나 여성 실험참가자들은 남성의 목소리보다 여성의 목소리를 유의미하게 더 잘 재인하였다(Wilding & Cook, 2000). 시각적 식별과 청각적 식별에서 조금은 다른 결과가 나왔는데 앞으로 더 다양한 연구를 통해 확인될 필요가 있다.

#### 확신도(Confidence)

목격자의 확신도는 그 증언이 일관적이든 비밀관적이든 상관없이 배심원의 판단에 강한 영향을 미치며(Brewer & Burke, 2002), 목격 증언의 신빙성을 판단하는데 있어 중요하게 고려되고 있다는 점에서 매우 중요한 요소로 고려된다(Manson v. Braithwaite, 1977; Neil v. Biggers, 1972). 하지만 시각적 목격자의 진술 정확도와 확신도에 대한 보편적인 믿음과는 달리 대부분의 과학적인 연구에서 확신도와 정확도의 관계가 상대적으로 약하거나 존재하지 않는 것을 확인하였다(Krug, 2007; Smith, Kassin, & Ellsworth, 1989). Kassin, Rigby와 Castillo(1991)에 따르면 회고적 자기 자각(Retrospective self-awareness: RSA) 조건에서 목격자의 정확도와 확신도 사이의 상관을 높여준다는 것을 발견하였다. RSA은 자신의 확신도를 평정하기 전에 자신이 식별을 결정했던 모습을 녹화한 비디오를 보여주는 것으로, 평가

를 내리는 시점의 자신의 모습을 다시 확인하도록 하는 것이 좀 더 객관적인 확신을 갖도록 할 수 있다.

이와 유사하게 음성식별에 있어서 익숙하지 않은 목소리에 대한 식별 정확성은 좋지 않으며 확신도와 정확도 사이에는 상관이 없는 것으로 나타났다(Olsson, Juslin, & Winman, 1998). Yarmey, Yarmey, Yarmey와 Parliament(2001)에 따르면 실험참가자들은 보통 수준의 익숙한 목소리보다는 매우 익숙한 목소리에 대해 유의미하게 더 확신했으며 익숙한 목소리일수록 더 높은 확신도를 나타냈다. 그러나 낮은 수준의 익숙함을 지닌 목소리와 익숙하지 않은 목소리 사이에는 확신도에 있어서 유의미한 차이가 나타나지 않았다. 이러한 결과는 음성식별로부터 얻어진 정보를 평가하는데 목격자의 확신도를 판단의 지표로 사용하는 것에 주의가 필요함을 시사한다.

#### 언어장막효과(Verbal overshadowing effect)

언어장막효과는 비언어적 자극을 언어적으로 묘사하게 했을 때 이후의 식별수행능력이 떨어지는 효과를 일컫는다(이윤, 조은경, 2011). 언어장막효과는 Schooler와 Engstler-Schooler (1990)에 의해 처음 확인된 이후 여러 학자들의 실험에 의해 반복되었다. Clare와 Lewandowsky(2004)는 대학생 참가자를 대상으로 언어장막효과에 대한 실험을 실시하였다. 범인에 대한 전체적인 언어적 표현 집단(holistic verbalization group)과 범인의 머리색, 이목구비에 대한 세부사항, 피부색 등의 특징적인 부분에 대한 언어적 표현 집단(featural verbalization group)의 범인식별 정확도를 분석한 결과, 언어적 표현을 한 두

집단은 식별 전 언어적 묘사를 실시하지 않았던 통제집단에 비해 더 낮은 식별 정확성을 보였다. 특징적인 부분에 대한 언어적 표현 집단과 통제집단을 비교하였을 때는 앞선 차이보다는 작은 차이가 발견되기는 하였지만 여전히 통제집단의 정식별률이 높았다. 이러한 언어장막효과는 아동에게 있어서도 동일하게 발견되었다(Dehon, Vanootighem, & Brédart, 2013)

이와 유사하게 범인의 음성에 대한 묘사는 이어지는 발화자 식별의 정확도를 감소시킬 수 있다(Yarmey, 2007). Perfect, Hunt와 Harris (2002)는 목소리 재인에서도 언어장막효과가 일어날 수 있음을 보여주었다. 대학생들을 대상으로 한 실험에서 음성식별에 앞서서 목소리를 묘사한 집단이 통제집단보다 좋지 않은 정확률을 나타냈다. Vanags 등(2005)의 연구에서는 같은 인종의 목소리(own-race voice) 또는 다른 인종의 목소리(other-race voice)의 재인에 있어서는 언어장막효과가 발생하지 않았으나 전화를 통해 목소리를 녹음하여 제시한 다른 실험에서 언어로 표현한 집단이 유의미하게 더 낮은 정식별률을 나타냈다.

#### 영국의 범인음성식별절차 지침

범인음성식별은 영국의 형사사법체계에서 찾아볼 수 있다. 1994년 발생한 강간사건에서 처음으로 음성식별절차에 대한 표준화된 절차의 중요성이 대두되었다. 이 사건에서 범인을 식별할 수 있는 주요 증거는 피해자가 범인의 목소리에 관해 진술한 것이었고 피해자는 경찰이 들려준 녹음된 목소리에서 한 남자의 목소리를 선택하였다. 당시 영국에는 범인음성

식별절차에 대한 공식지침이 없었기 때문에 시각적 범인식별절차의 공식지침을 따라 음성 식별절차를 진행하였다. 음성식별절차를 진행하는 과정에서 용의자의 음성 샘플은 신문과정에서 녹음된 음성을 편집하여 사용한 반면, 다른 목소리들은 편집되지 않은 독백을 사용하였는데 이후 재판 과정에서 용의자와 들러리들의 다른 음성 샘플 사용은 식별과정에 영향을 미칠 가능성이 있음이 전문가에 의해 제기되었다. 이후, 범인의 목소리에 대한 진술이 주요한 증거로 제시된 사건이 발생함에 따라 표준화된 범인음성식별절차의 중요성을 인식한 잉글랜드와 웨일즈에서는 비공식적인 지침을 만들어서 사용하였다. 그 후, 2003년 12월 잉글랜드와 웨일즈 정부에서는 범인음성식별절차의 공식지침인 'Advice on the use of voice identification parades'를 발행하게 된다(Wilcock et al., 2008). 범인음성식별절차에 대한 자료 준비과정부터 실시까지 전반적인 부분에 대해 다루고 있으며 제시되는 목소리의 수, 들러리의 특성, 전문가의 요건 등 상세하게 지침을 정해놓고 있다.

영국의 범인음성식별절차는 (1)재료(material) 준비, (2) 전문가 증인, (3) 오디오/목소리 절차 실행으로 구분된다(부록 참고). 첫 번째, 자료 준비는 식별절차에 사용할 음성재료를 준비하는 단계이다. 담당 경찰관은 음성식별을 위해서 목격자로부터 범인음성에 대한 상세한 진술을 얻어야 한다. 이때, Turnbull 가이드라인<sup>1)</sup>

1) Turnbull 가이드라인

- (1) 얼마나 오랫동안 목격자가 범인을 보았는가?
- (2) 목격자가 범인과 얼마나 떨어진 위치에 있었는가?
- (3) 범죄를 관찰하기 좋은 밝은 장소에서 보았는가?

은 음성식별에 있어서도 중요하게 고려된다. 목격자가 진술한 범인의 목소리에 대한 모든 내용은 범음성학/언어학 전문가에게 제공되며, 용의자의 목소리에 대한 목격자의 진술과 관련 묘사는 어떤 것이라도 식별절차 이전에 용의자/변호사에게 공개된다. 파지간격은 음성식별에 있어서 중요한 요소이다. 목격자의 기억력이 저하되거나 희미해질 수 있기 때문에 음성식별절차는 사건발생 4~6주 이내로 실시하는 것을 권장하고 있다. 또한 목소리가 스트레스에 의해서 영향을 받을 수 있기 때문에 자연스럽게 말하는 발화 샘플이 적절하며 발화 샘플은 용의자와 비슷한 배경(인종, 나이, 지역적·사회적 배경 등)의 사람들로부터 얻어야 한다.

음성식별절차에는 전문가 증인의 요건과 역할도 포함되어 있다. 전문가 증인은 국제 범음성학회(International Association of Forensic Phonetics)의 회원으로 음성학/언어학 분야에서 인증된 전문가 증인이어야 하며 전문가 증인은 라인업에서 제시할 목소리를 최종적으로 선택, 편집하고 용의자의 목소리를 가능한 정확하고 균형 있게 배치하는 역할을 하게 된다. 식별 담당 경찰관은 식별절차 전반에 걸쳐 재료의 보안과 보전에 대한 책임이 있다. 그리고 샘플들이 공평하게 구성되었음을 확실하게 하기 위해 들러리 샘플을 용의자의 것과 비교

- (4) 범죄를 관찰하는데 장애물이 있었는가?
- (5) 목격자는 피고인을 전에 본 적이 있는가?
- (6) 얼마나 자주 보았는가?
- (7) 자주 본 것이 아니라면, 피고인을 기억하는 특별한 이유가 있는가?
- (8) 최초의 목격과 범인식별 사이의 시간 간격은 얼마나 되는가?
- (9) 목격자가 최초에 경찰에 제시한 범인에 대한 묘사와 용의자의 외모에 현저한 차이가 있는가?

하여 말씨, 억양, 말의 빠르기 등을 조사해야 한다. 지침에서 전문가와 담당 경찰관은 모의 목격자를 활용한 여러 번의 실험을 거치는 것을 권장하고 있다. 실험은 모의 목격자들에게 사건의 간단한 개요를 제공하고 샘플을 들려준 다음, 용의자를 고르도록 하는 것으로, 용의자의 음성 샘플이 라인업에서 두드러져 다른 샘플보다 지목될 가능성이 높은지 아니면 라인업이 공평하게 구성되었는지를 판단할 수 있게 한다. 라인업에 제시되는 목소리의 길이는 각각 1분 정도여야 하며 개수는 용의자의 음성과 들러리의 음성 8개 즉, 9개의 음성을 제시하는 것으로 명시되어 있다. 식별 담당 경찰관은 목격자에게 라인업에 용의자가 있을 수도 있고 없을 수도 있다는 비편향 지시를 해주어야 하며 목격자가 원한다면 음성 샘플을 원하는 만큼 들려주어야 한다.

영국의 범인음성식별절차는 전반적으로 음성식별절차가 목격자와 용의자 모두에게 공정하고 공평한 절차여야 함을 강조한다. 용의자의 목소리에 대한 목격자의 진술과 ‘첫 묘사’는 식별절차 이전에 용의자/변호사에게 공개되어야 하며 본격적인 음성식별절차를 실행할 때 변호사가 참석할 기회를 주고 있다. 식별절차를 시작할 때 녹음테이프가 들어있는 밀봉된 가방은 목격자와 식별 담당 경찰관, 변호사의 입회하에 개봉 가능하다. 또한 변호사는 9개의 음성이 다른 순서로 녹음되어 있는 3개의 테이프 중에서 하나를 선택할 수 있는 기회를 갖는다. 이러한 식별절차과정은 녹화되어야 하며 그 절차가 끝난 후, 용의자는 그것을 볼 수 있는 기회를 갖게 된다. 용의자와 목격자 모두에게 한 쪽으로 치우치지 않는 공정한 절차를 진행하기 위해 용의자와 목격자 상황을 고려해야 한다.

## 우리나라 수사기관의 범인음성식별절차 도입을 위한 제언

현재 과학수사의 상당한 발전과 더불어 CCTV, DNA 등의 객관적인 증거는 사건을 해결하는데 있어서 역할이 확장되고 있다. 다만 여전히 강간 등 특정 범죄 유형의 경우 객관적인 증거가 존재하지 않아 피해자 혹은 목격자 진술이 사건의 해결에 있어서 유일한 증거가 되기도 한다. 이때 진술증거의 정확성 및 신빙성은 사건을 해결하는데 있어 중요한 역할을 한다. 인간의 기억에 기초한 진술은 부정확할 수 있으며 사건의 저장 및 인출 과정에서 외적인 요인에 의해 왜곡될 수 있다는 점을 고려할 때 진술 증거를 과학적이고 합리적이고 신뢰할 수 있는 방향으로 활용하는 방안을 모색하는 것은 상당히 중요하다. 지금까지 연구 결과들은 범인음성식별절차를 비롯하여 범인식별과정에서 영향을 미칠 수 있는 요인에 대해 제안하였다. 이를 바탕으로 영국 등의 나라에서는 표준화된 범인식별절차를 마련하여 진술 증거의 증명력을 향상시키고자 노력해 왔다. 우리나라의 경우, 범인음성식별을 비롯하여 범인식별과정에 영향을 미칠 수 있는 요인에 대한 연구는 거의 실시되지 않고 있으며 실무에서 활용 가능한 표준화된 가이드라인 또한 마련되어 있지 않은 실정이다. 이에 지금까지 연구 결과를 바탕으로 진술 증거의 정확성 및 객관성을 확보할 수 있는 표준화된 가이드라인의 개발은 시급하다고 볼 수 있다.

우리나라 수사기관의 범인음성식별절차 도입을 위한 제언은 다음과 같다. 첫째, 범인음성식별절차 도입을 위해서 다양한 인구통계학적 배경을 가진 사람으로부터 구성된 목소리

데이터베이스를 구축하는 것이 필요하다. 들러리들은 용의자와 같은 성별, 나이, 억양과 말투를 사용하는 사람이어야 하며 비슷한 사회적, 경제적, 교육적 배경을 가지고 있어야 한다(Broeders & Amelvoort, 1999; Hollien, 2012). 추가적으로 음성 제시 방법 및 비편향 지시문 등을 포함한 구조화된 지침은 식별절차의 객관성과 타당성을 확보하여 진술 증거의 증명력을 향상시키는데 도움을 줄 수 있다. 지금까지 연구 결과는 목격자 증언에 영향을 미칠 수 있는 요인으로 라인업 진행 방법을 가장 중요하게 고려하고 있다(Kassin, Ellsworth, & Smith, 1989). 특히 식별 담당자가 라인업 상의 용의자를 알고 있는 경우, 목격자에게 어떠한 암시나 영향을 미칠 수 있으므로 식별 담당 경찰관은 라인업 과정에서 관여해서는 안 된다(Malpass & Devine, 1981). 또한 목격한 사건에 대한 기억은 시간이 지남에 따라 흐려질 수 있으므로 사건 직후 식별은 최대한 빠르게 진행되어야 한다(Clifford, 1980; Valentine et al., 2003; Yarmey & Matthys, 1992; Yarmey, 2007). 식별절차에 영향을 미칠 수 있는 다양한 요인을 고려하여 우리나라 실정에 맞는 식별절차가 필요하다.

둘째, 합법적 절차의 도입과 더불어 일선 담당자들의 인식 개선은 우리나라 수사기관의 범인음성식별절차 도입에 있어 절대적으로 필요하다. 범인식별에 있어서 목격자와 용의자를 일대일로 대면하여 확인하게 했다면 사람의 기억의 한계와 무의식적인 왜곡 가능성을 비추어 지목 진술의 높은 신빙성을 부여할 수 없으며, 범인식별과정에서 라인업을 실시해야 함을 명시한 판례들은 경찰 및 검찰이 범인식별 관련 절차를 마련하는 계기가 되었다(대법원 2001.2.9. 선고 2000도4946 판결; 대법원

2002.9.10. 선고 2002도2947 판결; 대법원 2003.3.28. 선고 2002도6611 판결; 대법원 2004.2.27. 선고 2003도7033 판결). 경찰청 범인 식별절차 지침에 따르면 식별절차는 원칙적으로 2인 이상으로 구성하며 가급적 다수의 들러리를 포함해야 한다. 또한 식별을 위한 들러리들은 용의자와 인상착의가 유사한 사람으로 선별해야 한다. 목격자는 심리적으로 안정된 상태에서 식별절차를 진행해야 하며, 모든 절차는 기록하도록 권고하고 있다(이무선, 2014). 비록 시각적 범인식별절차에서는 지목진술의 신뢰성을 확보하기 위한 최소한의 규칙을 규정해 놓았으나 사실상 일선에서 지침이 제대로 지켜지지 않고 있는 실정이다(권순민, 2009; 김지영, 김시업, 2006; 이성기, 2010). 경찰들은 범인식별을 위해 라인업을 실시하는 경우는 매우 드물었으며, 일반적으로 용의자와 몇 명의 들러리들의 사진을 제시하는 경우가 대부분이었다(김지영, 김시업, 2006; 이성기, 2010). 이때 들러리 사진 역시 경찰 내부 전산망에 보관된 전과자 사진이나 주민등록증, 운전면허증 등 각종 신분증 사진 또는 SNS 등을 통해 얻는 방식으로 일관성이 없어 식별과정에 영향을 미칠 수 있는 요인이 포함되어 있으며, 사실상 라인업이나 사진 제시 등의 방식을 사용하지 않은 채 직접 대면하도록 하는 경우도 상당 수 포함하고 있는 것으로 나타났다(이성기, 2010). 범인식별절차가 수사과정에서 안정적으로 정착하기 위해서는 일선 담당자들의 인식의 변화가 선행되어야 하지만 여전히 식별절차에 대한 인식은 확립되지 못한 것으로 보인다. 담당 경찰관들에 대한 설문조사에서 범인식별에 있어서 라인업을 실시한다고 응답한 수는 5.5%에 불과하였으며 실시하지 않은 이유에 대해서는 인

원확보 곤란(44%), 범인식별에 도움이 안됨(14.5%), 번거로움(39.4%), 협소한 장소(2%)로 응답하여 응답자 중 절반가량의 경찰관이 라인업은 식별에 도움이 되지 않거나 번거롭다와 같이 부정적인 이유로 라인업을 실시하지 않은 것으로 나타났다(이성기, 2010). 추가적으로 범인식별절차의 문제점으로는 절반가량의 경찰관들이 목격자의 비협조를 들었으며, 라인업의 인원 및 예산확보, 전문성 부족 등을 언급하였다.

기존 판례에서는 지목 진술의 증명력을 확보하기 위해서 반드시 식별 지침을 따를 것을 강조하고 있으며, 적법하지 않은 절차의 사용으로 인해 증거효력을 잃게 되는 경우가 발생할 수 있음에도 일선에서는 여전히 이에 대한 관심과 필요성에 대한 인식이 부족한 것으로 보인다. 우리나라 형사소송법에서는 기본적으로 피의자 또는 피고인의 기본권을 침해하지 않으며 공정한 수사과 재판 절차를 제공 받을 수 있도록 권리를 보장하고 있다(예를 들어, 형사소송법 제244조의3(진술거부권 등의 고지), 변호사선임권 제30조(변호사선임권자) 등). 범인음성식별절차의 제도적 도입 역시 피의자 또는 피고인에게 공정한 절차를 제공하고자 하는 동일한 목적도 지니고 있다. 피해자 또한 적절하지 않은 절차로 인해 증거 채택이 거부되어 불이익을 받는 경우가 없어야 할 것이며 조사 과정에서 피의자를 직접 대면하게 하는 등의 절차로 인한 2차 피해를 줄일 수 있도록 해야 할 것이다.

마지막은 지속적인 연구의 필요성이다. 범인음성식별절차는 사람에 의해 행해지는 것으로 다양한 오류가 발생할 수 있다. 예를 들어, 목소리의 친숙함(familiarity), 전화를 통한 범인의 목소리, 무기집중효과 등은 음성식별절차

에 영향을 미칠 수 있으나 이에 대한 연구는 여전히 부족한 실정이다. 음성식별절차에 영향을 미칠 수 있는 다양한 요인에 대한 기초 연구들은 목격자 음성식별의 정확성을 향상하는데 커다란 기여를 할 수 있다. 이와 더불어 우리나라 형사사법체계에 적용할 수 있는 과학적이고 공정한 범인음성식별절차에 대한 연구를 지속적으로 진행하여 우리나라 실정에 적합하게 목격자의 범인음성에 대한 진술의 정확도를 높일 수 있는 방안을 마련해야 할 것이다.

본 연구는 음성식별절차에 대해 소개하였다. 아직 우리나라에서 범인음성식별과 관련된 인식은 미미한 수준이며 여전히 많은 연구와 논의가 이루어져야 할 분야이다. 범인의 음성에 관한 진술이 신빙성을 확보하기 위해서는 지속적인 연구와 더불어 범인음성식별절차가 제대로 정착되고 실시될 수 있도록 관계자들의 관심과 관리가 필요할 것이다.

## 참고문헌

- 권순민 (2009). 라인업 절차의 합리적 운영 방안에 대한 연구. *형사법연구*, 21, 429-458.
- 김지영, 김시업 (2006). 목격자 증언의 정확성 제고방안. *연구총서*, 2006(23), 1-136.
- 대법원 1984.6.26. 선고 84도947 판결
- 대법원 2001.2.9. 선고 2000도4946 판결
- 대법원 2002.9.10. 선고 2002도2947 판결
- 대법원 2003.3.28. 선고 2002도6611 판결
- 대법원 2004.2.27. 선고 2003도7033 판결
- 대전고등법원 2008.6.18. 선고 2008노50 판결
- 서울고등법원 1998.7.8. 선고 97노1260 판결
- 이무선 (2014). 경찰수사에서 범인식별에 관한

- 목격자진술의 신빙성 제고 방안. *경찰학연구*, 14(3), 3-32.
- 이성기 (2010). 목격자의 범인식별진술의 증명력을 높이기 위한 실질적 대안으로서의 비디오 라인업. *경찰학연구*, 10(2), 117-142.
- 이 윤, 조은경 (2011). 범인음성식별절차에 있어서의 언어장막효과(Verbal Overshadowing Effect)와 눈감기 효과. *한국심리학회지: 사회 및 성격*, 25(4), 77-92.
- 홍수기 (2006). 음성 개인식별. *물리학과첨단기술*, 제15권 제4호, 12-19.
- ADVICE ON THE USE OF VOICE IDENTIFICATION PARADES (2003). <http://webarchive.nationalarchives.gov.uk/20130125102358/http://www.homeoffice.gov.uk/about-us/corporate-publications-strategy/home-office-circulars/circulars-2003/057-2003/>에서 2015년 10월 1일 자료 얻음.
- Bedau, H. A., & Radelet, M. L. (1987). Miscarriages of justice in potentially capital cases. *Stanford Law Review*, 21-179.
- Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior*, 26(3), 353.
- Broeders, A. P. A., & Van Amelsvoort, A. G. (1999). Lineup construction for forensic earwitness identification: a practical approach. In *Proceedings of the 14th International Congress of Phonetic Sciences* (Vol. 2, pp. 1373-1376). San Francisco, CA.
- Bull, R., & Clifford, B. R. (1984). Earwitness voice recognition accuracy. *Eyewitness testimony: Psychological perspectives*, 92-123.
- Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic bulletin & review*, 7(1), 26-48.
- Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 739.
- Clifford, B. R. (1980). Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior*, 4(4), 373.
- Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior*, 11(3), 233.
- Dehon, H., Vanootighem, V., & Brédart, S. (2013). Verbal overshadowing of face memory does occur in children too!. *Frontiers in psychology*, 4.
- Hollien, H., Majewski, W., & Hollien, P. A. (1974). Perceptual identification of voices under normal, stress, and disguised speaking conditions. *The Journal of the Acoustical Society of America*, 56(S1), S53-S53.
- Hollien, H. (2012). On earwitness lineups. *Investigative Sciences Journal*, 4(1).
- Huff, C. R., Rattner, A., Sagarin, E., & MacNamara, D. E. (1986). Guilty until proved innocent: Wrongful conviction and public policy. *Crime & Delinquency*, 32(4), 518-544.
- Kassin, S. M., Rigby, S., & Castillo, S. R. (1991).



- The accuracy-confidence correlation in eyewitness testimony: Limits and extensions of the retrospective self-awareness effect. *Journal of Personality and Social Psychology*, 61(5), 698.
- Kassin, S. M., & Neumann, K. (1997). On the power of confession evidence: an experimental test of the fundamental difference hypothesis. *Law and human behavior*, 21(5), 469.
- Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The “general acceptance” of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist*, 44(8), 1089.
- Kersta, L. G. (1962). Voiceprint identification. *The Journal of the Acoustical Society of America*, 34(5), 725-725.
- Kerstholt, J. H., Jansen, N. J., Van Amelsvoort, A. G., & Broeders, A. P. A. (2004). Earwitnesses: effects of speech duration, retention interval and acoustic environment. *Applied cognitive psychology*, 18(3), 327-336.
- KNELLER, W., MEMON, A., & STEVENAGE, S. (2001). Simultaneous and Sequential Lineups: Decision Processes of Accurate and Inaccurate Eyewitnesses.
- Krug, K. (2007). THE RELATIONSHIP BETWEEN CONFIDENCE AND ACCURACY: CURRENT THOUGHTS OF THE LITERATURE AND A NEW AREA OF RE-SEARCH. *Applied Psychology in Criminal Justice*, 3(1).
- Lindsay, R. C. L., & Wells, G. L. (1985). Improving Eyewitness Identifications From Lineups: Simultaneous Versus Sequential Lineup Presentation. *Journal of Applied Psychology*, 70(3), 556-564.
- Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about “weapon focus”. *Law and Human Behavior*, 11(1), 55.
- Luus, C. A., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior*, 15(1), 43.
- Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of applied Psychology*, 66(4), 482.
- Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of experimental child psychology*, 27(1), 153-165.
- Manson v. Braithwaite, 432 U.S. 98 (1977).
- McGrath, M., & Turvey, B. E. (2014). Eyewitness Identification: Uncertainty, Error, and Miscarriages of Justice. *Miscarriages of Justice: Actual Innocence, Forensic Evidence, and the Law*, 91.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, 7(1), 3.
- Milne, R., & Bull, R. (1999). *Investigative interviewing: Psychology and Practice*. Chichester: John Wiley.
- Morgan, C. A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., & Southwick, S. M. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International journal of law and psychiatry*, 27(3), 265-279.

- Narby, D. J., Cutler, B. L., & Penrod, S. D. (1996). The effects of witness, target, and situational factors on eyewitness identifications. *Journal of Applied Psychology, 75*(3), 358.
- Neil v. Biggers 409 U.S. 188 (1972).
- Nosworthy, G. J., & Lindsay, R. C. (1990). Does nominal lineup size matter?. *Journal of Applied Psychology, 75*(3), 358.
- Öhman, L., Eriksson, A., & Granhag, P. A. (2010). Overhearing the Planning of A Crime: Do Adults Outperform Children As Earwitnesses?.
- Olsson, N., Juslin, P., & Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. *Journal of Experimental Psychology: Applied, 4*(2), 101.
- Orchard, T. L., & Yarmey, A. D. (1995). The effects of whispers, voice sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology, 9*(3), 249-260.
- Patterson, K. E., & Baddeley, A. D. (1977). When Face Recognition Fails. *Journal of Experimental Psychology, 3*(4), 406-417.
- Perfect, T. J., Hunt, L. J., & Harris, C. M. (2002). Verbal overshadowing in voice recognition. *Applied Cognitive Psychology, 16*(8), 973-980.
- Pozzulo, J. D., & Lindsay, R. C. L. (1998). Identification accuracy of children versus adults: a meta-analysis. *Law and Human Behavior, 22*(5), 549.
- Pozzulo, J. D., & Marciniak, S. (2006). Comparing identification procedures when the perpetrator has changed appearance. *Psychology, Crime & Law, 12*(4), 429-438.
- Procter, E. E., & Yarmey, A. D. (2003). The Effect of Distributed Learning on the Identification of Normal-Tone and Whispered Voices. *The International Journal of Creativity & Problem Solving, 13*(1), 17-29.
- Rattner, A. (1988). Convicted but innocent: Wrongful conviction and the criminal justice system. *Law and Human Behavior, 12*(3), 283.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive psychology, 22*(1), 36-71.
- Searcy, J. H., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory & cognition, 27*(3), 538-552.
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*(2), 139.
- Smith, V. L., Kassin, S. M., & Ellsworth, P. C. (1989). Eyewitness accuracy and confidence: within-versus between-subjects correlations. *Journal of Applied Psychology, 74*(2), 356.
- Stebly, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21*(3), 283.
- Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17*(8), 969-993.
- Valentine, T., & Mesout, J. (2009). Eyewitness identification under stress in the London Dungeon. *Applied Cognitive Psychology, 23*(2),

- 151-161.
- Vanags, T., Carroll, M., & Perfect, T. J. (2005). Verbal overshadowing: A sound theory in voice recognition?. *Applied Cognitive Psychology*, 19(9), 1127-1144.
- Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, 36(12), 1546.
- Wilcock, R., Bull, R., & Milne, B. (2008). *Witness identification in criminal cases: Psychology and practice*. Oxford University Press.
- Wilding, J., & Cook, S. (2000). Sex differences and individual consistency in voice identification. *Perceptual and Motor skills*, 91(2), 535-538.
- Yarmey, A. D., & Matthys, E. (1992). Voice identification of an abductor. *Applied Cognitive Psychology*, 6(5), 367-377.
- Yarmey, A. D., Yarmey, A. L., Yarmey, M. J., & Parliament, L. (2001). Commonsense beliefs and the identification of familiar voices. *Applied Cognitive Psychology*, 15(3), 283-299.
- Yarmey, A. D. (2007). *The psychology of speaker identification and earwitness memory* (pp. 101-136). Mahwah, NJ: Lawrence Erlbaum Associates
- 1 차원고접수 : 2016. 06. 25.  
심사통과접수 : 2016. 07. 11.  
최종원고접수 : 2016. 07. 21.

## Introducing voice identification in a criminal investigation and its implications for legal policy in South Korea

Minseon Bae

Department of Psychology  
Hallym University

Misun Yi

Institute of Psychology and Law  
Hallym University

Voice identification is the process of comparing a perpetrator's and suspect's voice samples in order to determine if the two voices are consistent with one another. Previous research suggested that since it requires remembering and recognizing voices from a past experience, the accuracy of an earwitness's identification could be affected by various factors. These include variables relating to the identification procedures such as line-up size, instruction, delay and so on. Also a witness's age, gender, race or confidence could affect the accuracy of voice identification results. Recognizing how vulnerable the identification process should be, the governments of England and Wales published the official guidance on voice parades in 2003. However, theoretical findings on voice identification and its application for the legal system in South Korea are still little known. These findings are required to continue with concerns and efforts from both academic and legal experts in order to establish the standardized guidance for voice identification in South Korea.

*Key words* : voice identification, voice line-ups/parades, earwitness, eyewitnesses identification, eyewitness testimony

## 부 록

### 영국의 범인음성식별절차 지침(ADVICE ON THE USE OF VOICE IDENTIFICATION PARADES)

#### 가) 재료(material) 준비

1. 음성식별을 위해서 담당 경찰관은 목격자로부터 목소리에 대한 자세한 묘사를 포함한 상세한 진술을 얻어야 한다. 특히 Turnbull 가이드라인은 목격자 범인식별진술 신빙성 판단 시 고려해야 하는 요인으로 음성식별에 있어서도 반드시 고려해야 한다. 범음성학/언어학 전문가에게 제공되는 자료에는 목격자가 진술한 범인의 목소리에 대한 모든 것이 반드시 포함되어야 한다. 목격자의 진술과 용의자 목소리에 대한 ‘첫 묘사’는 어떤 것이라도 식별절차 진행되기 이전에 용의자/변호사에게 공개해야 한다.

2. 어떠한 경우라도 라이브 용의자 및 들러리를 사용하여 라이브 음성 인식 절차를 수행하여야 한다.

3. 식별 담당 경찰관은 용의자의 목소리를 대표하는 샘플을 얻어야 한다. 적절한 공급원은 질문에 답을 하면서 자연스럽게 말하는 용의자를 녹화한 경찰 인터뷰 테이프이다(전문가들은 목소리가 스트레스에 의해 영향을 받을 수 있다고 조언하였음). 용의자와 인터뷰를 시작할 때 녹음된 인터뷰 샘플은 식별 목적을 위해 사용될 수 있다고 알려주고 동의를 구한다.

4. 식별 담당 경찰관은 용의자의 지역적, 사회적 배경과 나이, 인종과 비슷한 사람으로부터 20개 이상의 발화 샘플을 얻어야 한다. 본 사건과 관련 없는 사건들로부터의 다른 경찰 인터뷰 녹화 테이프가 적절한 공급원이 될 수 있다.

5. 식별 담당 경찰관은 적당한 시간 안에 확실하게 모든 일을 수행하고 완료해야 한다. 절차는 사건 발생 후 4-6 주 이내에 수행되는 것이 좋다. 목격자의 측면에서 기억력의 저하되거나 회미해지는 것이 이 분야의 전문가에 의해서 중요한 요소로 밝혀져 왔기 때문이다.

6. 식별 담당 경찰관은 샘플 목소리의 최종적인 선택과 편집을 확실하게하기 위해서 그리고 용의자의 목소리와 가능한 정확하고 균형 있게 매치하기 위해서 International Association of Forensic Phonetics의 회원인 음성학/언어학 분야에서 인증된 전문가 증인을 요청해야 한다.

## 나) 전문가 증인

1. 용의자의 목소리와 비슷한 목소리들과 함께 용의자의 목소리 샘플을 포함하는 테이프는 임명된 전문가 증인에게 보내져야 한다. 식별 담당 경찰관은 용의자의 테이프를 명확하게 표시해야 한다. 나머지 테이프들은 성(姓)을 표시하거나 관리 조회번호(custody reference number)를 표시해야 한다.

2. 전문가는 한 묶음의 테이프들로부터 발화 샘플을 고르는 일을 하게 된다. 각각 1분 분량이어야 하고 말하기의 다양한 단편들 그리고/또는 연속적인 말하기로 구성되어야 한다. 각 샘플이 다른 단어나 주제를 포함하는지는 상관없다. 모두 9개의 샘플이 선택되어야 한다(즉, 용의자의 목소리+들러리 8개).

3. 총 9개의 말하기 샘플은 3개의 비디오카세트에 녹음이 되어야 한다. 각각은 다른, 무작위의 순서로 녹음된다. 샘플들은 매겨진 번호와 함께 제시된다. 식별 담당 경찰관들은 각 비디오에 대한 색인을 준비해야 하며 각 샘플의 이름/참고정보를 설명하고 번호를 할당한다. 3개의 비디오는 명확하게 A, B, C로 표시해 준비되어야 하고 각 샘플에 대한 조회 번호는 해당 샘플이 생성되는 동안 화면을 통해 제시되어야 한다. 각 테이프 샘플은 3번 반복된다.

4. 식별 담당 경찰관은 선택된 샘플이 어떤 개인이나 범죄자로서 특정 샘플을 식별을 하게 한다는 의문을 제기할 만한 것이 없는 정당한 것임을 보장할 책임이 있다. 8명의 들러리 샘플 각각은 용의자의 것과 비교하여 공정하다는 것을 확실히 하기 위해 말씨, 억양, 음의 높이, 톤, 말하는 빠르기를 조사해야 한다.

5. 전문가와 식별 담당 경찰관이 모의 목격자를 활용해서 다수의 테스트를 하는 것을 강력하게 권장한다. 모의 목격자들에게 사건의 간단한 개요를 제공해야 한다. 그들은 통제된 조건에서 일련의 샘플을 듣는 것을 요청받고 그 범죄에 대한 용의자를 고르도록 요청받게 된다.

6. 테스트의 결과에 대한 모든 샘플들의 추후 조사는 다음을 확실하게 한다. 말하는 것을 포함하고 있지 않다면 하나 또는 다른 샘플이 용의자의 것이라는 불공정한 가정으로 이끌 것이다. 그 말투가 없다면 하나 또는 다른 샘플이 용의자의 것이라는 불공정한 가정으로 이끌 것이다.

7. 이러한 실험 결과는 수행되어 온 절차가 객관적이고 신중하고 균형 잡힌 방법임을 설명하는 전문가 증인에 의해 제공되는 증거의 일부로서 형태를 갖추어야 한다.

8. 선택된 9개의 오디오 테이프는 하나의 봉투에 봉해야 하고 봉투에 색인을 하고 별개의 밀

봉된 봉투에 넣어야 한다. 재료의 준비와 관련하여 수행된 작업을 설명하면서 각각의 가방에 서명하고 날짜를 적는다. 그리고 전문가 증인의 진술이 준비된다. 전문가는 식별 담당 경찰관에게 밀봉된 가방으로 모든 완성된 재료를 제시해야 한다.

9. 3개의 샘플 비디오테이프와 관련된 색인의 준비가 완료되면 이것은 그 작업을 수행하는 임명된 전문가에 의해서 경찰 증거 봉투에 밀봉되어야 한다.

10. 식별 담당 경찰관은 식별 절차 과정 전반에 걸쳐 재료의 보안과 보전에 대한 책임이 있다.

11. 이러한 절차들은 용의자에게 들리리들의 샘플에 대해 검토하거나 거부할 어떤 기회도 제공하지 않는다.

#### 다) 오디오/목소리 절차 실행

1. 용의자의 변호사는 목소리 식별 절차 실행할 때, 참석할 기회가 주어진다. 테이프가 있는 밀봉된 가방은 목격자와 식별 담당 경찰관이 참석한다면 변호사 입회하에 개봉 가능하다.

2. 식별 절차는 비디오 녹화되어야 하고 이후, 용의자는 적절한 때에 그것을 볼 기회를 갖는다.

3. 변호사는 재생될 샘플테이프를 선택할 기회를 갖는다(즉, A, B, 또는 C).

4. 식별 담당 경찰관은 증인에게 절차 중에 재생되는 목소리 샘플 중에는 용의자의 목소리가 있을 수도 있고 없을 수도 있다는 점을 설명해주어야 한다. 식별 담당 경찰관은 증인에게 선택하기 전에 각 테이프를 적어도 한 번 이상 듣도록 지시를 한다. 증인은 어떤 샘플이나 모든 샘플을 그들이 원하는 만큼 들을 수 있다.

5. 식별 담당 경찰은 목격자가 언급한 내용이나 선택을 모두 기록한다.

6. 절차에 따라서 목격자로부터 진술을 얻어야 하고 식별 절차와 그들의 선택에 대해 기록한다. 증인이 식별 절차를 수행한 방을 떠나면 비디오테이프는 VCR 기계에 남아서 돌아가고 있어야 한다. 식별 담당 경찰만이 테이프와 관련된 색인을 포함하는 밀봉된 가방과 봉투를 열 수 있으며, 변호사에게는 나타난 세부사항들을 기록할 기회가 제공된다.

7. 식별 담당 경찰은 재판을 위해 음성식별절차와 관련 있는 모든 재료를 보관한다.