

국립국어원 서울말 낭독체 발화 말뭉치 (버전 2.0)

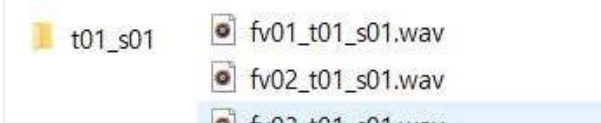
- **자료명:** 국립국어원 서울말 낭독체 발화 말뭉치
- **공개일:** 2021. 3. 30 (버전 2.0)
 - ※ 버전 2.0에서 수정, 보완된 사항
 - 낭독자가 대본과 다르게 읽은 경우 그 부분을 정서법에 맞게 전사하여 대본을 수정
 - 대본과 불일치한 음성 파일을 대본에 맞게 재분할하여 수정(305건)
- **자료 유형:** 음성, 텍스트
- **자료 설명**
 - ※ 본 자료는 연구용으로만 활용할 수 있습니다.
 - 내용
 - 원본은 '03년 서울말의 변화를 조사할 목적으로 930개 문장 분량의 대본을 118명이 서울말로 읽은 음성과 대본 자료임
 - 원본을 대본과 음성 파일이 일치하도록 수정, 보완하였음
 - 제보자들의 발화는 Digital Audio Tape와 IBM 호환 PC에 동시에 녹음
 - 녹음에 사용된 마이크는 미국 SHURE사의 다이내믹 마이크인 SM48 또는 SM58
 - A/D 변환은 미국 Syntrillium Software Co.의 CoolEdit 2000 ver 1.0으로 16bit,16kHz 양자화 및 표본 추출
 - 분량 및 형식
 - 음성 파일: 문장 단위 음성 파일(wav 파일 총 87,035개, 16.5 GB)
 - 수정 대본: 음성 파일별로 수정 대본을 연계한 자료(xlsx 파일 1개, 4.71MB)
 - 문서 목록 : 문서 기호 및 문서명(xlsx 파일 1개, 9.80KB)
 - 낭독자 세부 정보: 낭독자별 연령, 성별 등 세부 정보(xlsx 파일 1개, 19.5KB)
 - 최초 대본: 최초 낭독자에게 제공한 대본(hwp 파일 1개, 52.2.KB)
 - 배포 파일 수 및 크기: 압축 파일(zip) 1개, 11.6GB
 - 인용
(국문) 국립국어원(2021). 국립국어원 서울말 낭독체 발화 말뭉치(버전 2.0). URL: <https://kli.korean.go.kr/corpus>

(영문) National Institute of Korean Language (2021). NIKL The Corpus of Reading Aloud Korean Short Stories by Seoul People (v.2.0). URL: <https://kli.korean.go.kr/corpus>

· 세부 설명

가. 음성 파일

- 음성 파일명은 ‘낭독자 기호’, ‘문서 기호’, ‘문장 기호’ 순으로 구성됨
예) fv01_t01_s01: 낭독자 fv01이 t01 문서의 첫 번째 문장을 낭독한 음성 파일
- 문장별 폴더 내 낭독자들이 각 문장을 읽은 음성 파일이 저장되어 있음



나. 수정 대본

- 음성 파일별 수정 대본의 문장을 연계
- ‘낭독자 구분’과 ‘문서 구분’은 각각 ‘낭독자 세부 정보’, ‘문서 목록’과 연계
- ‘단락 구분’은 단락의 순서임(예: t01_p01은 t01 문서의 첫 번째 단락)
- ‘문장 구분’은 문장의 순서임(예: t01_s01은 t01 문서의 첫 번째 문장)
- ‘음질 구분’은 ‘정상’, ‘음질 불량’으로 구분
- ‘대본과의 비교’는 낭독자가 최초 대본대로 읽었는지 여부이며 ‘일치’, ‘불일치’로 구분
- 예시

음성 파일	수정 대본	낭독자 구분	문서 구분	단락 구분	문장 구분	음질 구분	대본과의 비교	최초 대본
fv01_t01_s01	기차도 전기도 없었다.	fv01	t01	t01_p01	t01_s01	정상	일치	기차도 전기도 없었다.
fv01_t01_s02	라디오도 영화도 몰랐다.	fv01	t01	t01_p01	t01_s02	정상	일치	라디오도 영화도 몰랐다.

다. 문서 목록 예시

문서 구분	문서명
t01	요람기
t02	소나기
t03	아름다움에 대하여
t04	방망이 깎던 노인
t05	문학의 세계와 삶의 세계
t06	독서와 인생
t07	수난 이대(受難二代)
t08	메밀꽃 필 무렵
t09	토끼와 자라
t10	선녀와 나무꾼
t11	호랑이와 곶감
t12	해님 달님
t13	그리운 시냇가
t14	광화문 지하도 아주머니

t15	막 지은 밥
t16	눈 오던 날
t17	송농의 지혜
t18	까만 눈동자들 앞에서
t19	내 고향 개울 물에서

라. 낭독자 세부 정보 예시

낭독자 구분	녹음년도	성별	녹음시 나이	출생지	성장지	부친 출신지	모친 출신지	녹음지	발성환경	잡음환경	녹취장비	A/D장비	통신종류
mv01	2003	남	만24	서울	서울	서울	서울	아주대	스튜디오	잡음배제	슈어sm48	pc	유선
mv02	2003	남	만23	서울	경기 수원	서울	서울	아주대	스튜디오	잡음배제	슈어sm48	pc	유선

- 자료 내용 문의: 02-2669-9638