FluencyBank: a briefing sheet for working with IRBs

FluencyBank is funded by both the US NIH (NIDCD) and the US National Science Foundation. Together with three other TalkBank databases (CHILDES, PhonBank,and AphasiaBank), it is included at <https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html> in the list of officially-sanctioned NIH Data Sharing Repositories. It is the newest of these four repositories

A large motivation for the construction of a new TalkBank (talkbank.org) curated FluencyBank database was that, unlike other areas of language study, speech fluency has notorious and well-documented transcription reliability issues. Therefore, simple transcripts, which form a proportion of TalkBank archives, are not best suited to the study of speech fluency. Speech fluency archives require either video or audio media for most research uses. However, inclusion of audio and particularly video requires careful attention to Human Subjects confidentiality issues.

In regards to confidentiality, the basic premises of FluencyBank ([www.talkbank.org/fluencybank](http://www.talkbank.org/fluencybank)) are:

1. All data in all TalkBank archives are fully de-identified. New recordings avoid the inclusion of identifying language. In older data, surnames and addresses are replaced by the capitalized terms “Lastname” and “Address”. Locations in the audio with such identifiers are replaced with silence.
2. All FluencyBank media linked data are password protected. Passwords may only be obtained by joining as a FluencyBank member. Only established researchers can become members. Students can only access data under the supervision of their faculty advisors. To join and receive the password, researchers must provide written justification for their interests to Brian MacWhinney at CMU. Passwords are changed at regular intervals.
3. Individual IRBs will be able to specify higher levels of protection, such as another level of password protection, or signed agreements for data use.

The NIH data-sharing guidelines that can be reviewed at <https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm> data state that, “in most instances, sharing and archiving of data is possible without compromising confidentiality and privacy rights. The procedures adopted to share data while protecting privacy should be individually tailored to the specific dataset.” In accord with this policy, the TalkBank repositories allow for further procedures for data protection as described at <http://talkbank.org/share/irb/options.html>

Many TalkBank datasets include video linked to transcripts. For example, AphasiaBank has password-protected video data from 464 participants, all of whom have provided informed consent. Collection and dissemination of these data has been uniformly approved by IRBs at every institution and has never been denied. IRBs have also approved data-sharing for over 100 data sets collected before web-based data-sharing was possible. Many of these include video linked to transcripts; others include audio linked to transcripts.

Video data are preferred for FluencyBank contributions because stuttering and other forms of speech disfluency are best disambiguated using visual signals (e.g., the difference between hesitation and vocal blocks). Additionally, the distinctions among normal, language-learning and stuttering disfluency include the presence of so-called secondary behaviors in stuttering, typically quantified for publication in subject severity descriptions. These are only available in the visual record (body movements, eye gaze deviations, struggle behaviors, etc.). However, if your IRB will not approve data-sharing of older video, we suggest requesting inclusion of the extracted audio in the database and archiving of the video offline in TalkBank backup media for data preservation.

Summary:

IRBs should approve all requests to contribute new data to FluencyBank, including transcripts, audio, and video, as long as participants provided informed consent for release of password-protected data. They should approve contributions of de-identified older audio and transcript data, depending on the language in the original informed consent forms. When in doubt regarding how to interpret the original language within the framework of modern technology, please consult us for advice. In some cases, IRBs will approve sharing of older video data. In other cases, they will only permit offline archiving of these data.