Fw: About data collection

Dr.Suryakanth V Gangashetty <svg@kluniversity.in>
Wed 1/25/2023 2:32 PM
To: Dr.Samudra Vijaya K <samudravijaya@kluniversity.in>;Dr GOVIND DIVAKARAN <d_govind@kluniversity.in>;Akhilesh Kumar Dubey <dubey18oct@kluniversity.in>;DLS <dls2science@gmail.com>;Rithik Sarab <190030004@kluniversity.in>;oindrila banerjee <195031002@kluniversity.in>
Cc: Dr.Suryakanth V Gangashetty <svg@kluniversity.in>

Dear All
    Kindly see the forwarding email from Hema A Murthy madam regarding suggested sampling frequency for the speech data collection by vendors.

Thank you
Yours Sincerely
Suryakanth V Gangashetty

From: Hema A. Murthy <hema@cse.iitm.ac.in>
Sent: Wednesday, January 25, 2023 8:25 AM
To: Prof. Umesh S <umeshs@ee.iitm.ac.in>
Cc: Dr.Suryakanth V Gangashetty <svg@kluniversity.in>; Dr.Samudra Vijaya K <samudravijaya@kluniversity.in>; Jeyanthi Arumugam <kumarajeyanthi@gmail.com>
Subject: Re: About data collection

Yes, this is very important.
Wideband: 16K or multiples thereof.
Narrowband: 8K PCM or mulaw or alaw?

On Tue, Jan 24, 2023 at 12:29 PM Prof. Umesh S <umeshs@ee.iitm.ac.in> wrote:
> Oops.. I forogt the most important
>
> proper sampling and storing format for both Wide-band and Telephone speech (uniform across vendors)
>
> On Tue, Jan 24, 2023 at 12:28 PM Prof. Umesh S <umeshs@ee.iitm.ac.in> wrote:
>> Dear Surya
>>
>> Thanks to you and team for the good work to start SQC.
>>
>> I have few comments, which is the basis on which we thought of data collection.
>> (1) We want to cover mostly monolingual speech, covering at least 4 dialects
>> (2) We want mostly conversational speech (part of which can be monologues) but mostly conversational (at least 50-60%).. monologues can be 20-30%The spontaneous speech should by 80% and remaining 15-20% can even be read speech to cover vocabulary etc
>> (3) Telephone speech should cover about 30% of entire data
>> (4) Transcription should be of good quality, and consistent ACROSS VENDORS
>> (5) For conversational, speaker markings should be helpful
>>
>> The remaining aspects, we can be flexible.. we will try and have maximum uniformity across vendors.. but small deviations. let us let go off..
>>
>> Thanks!
>>
>> Umesh