

# SphereDiar - Data tables

Tuomas Kaseva

April 2019

## 1 Meeting corpus

Table 1: Removed meetings

AMI	EN2001a, EN2001e, EN2002c, EN2003a, EN2006a EN2006b, IB4005, IS1003b
ICSI	Bmr012

Table 2: *ICSI<sub>eval</sub>*

Bed017, Bmr014, Bed009, Bro017, Bsr001, Buw001, Bmr003, Bro024, Bns002
---

Table 3: *Clust<sub>dev</sub>*

IS1003d, ES2016b, ES2010a, ES2006a, TS3012b, TS3008b, TS3011d, Bro008, Bed013, Bed016, Bmr022, Btr002
--

Table 4: Frame categorization

	Percentage (%)	Number of frames
$H_{\geq 65\%}$	84.2	595 086
$H_{< 65\%}$	15.8	111 592
Total	100	706 678

## 2 Speaker corpora

Table 5: Gender distribution

	Number of females	Number of males	Number of speakers
$LS_{1000}$	500	500	1000
$LS_{2000}$	987	1013	2000
$VC_{1000}$	500	500	1000
$VC_{2000}$	731	1269	2000

Table 6: Frame compositions

	Minimum number of frames per speaker	Maximum number of frames per speaker	Number of frames
$LS_{1000}$	382	764	654 297
$LS_{2000}$	341	670	1 204 967
$VC_{1000}$	838	1000	995 443
$VC_{2000}$	577	670	1 337 601