Comparing Human- and Machine-Annotated Language Input Across Childcare Settings

DAYCARE VS. HOME

MELANIE SODERSTROM
WAYNE FRANZ
UNIVERSITY OF MANITOBA
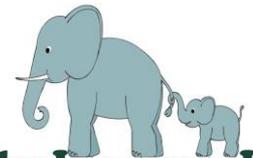
UNIVERSITY OF MANITOBA

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

Baby Language Lab

Toward largescale annotated child language data

- LENA

- Is it as good as human annotation?
  - NO! (Of course not)

- How good is it?

- What are its strengths and weaknesses?

How good is LENA?

- AWC
  - r = .92 LENA Technical Reports
  - r = .64 (Canault et al. 2015, French)
  - r = .80 (Weisleder & Fernald, 2013, Spanish)
  - r = .93 (Caskey et al. 2014, NICU)
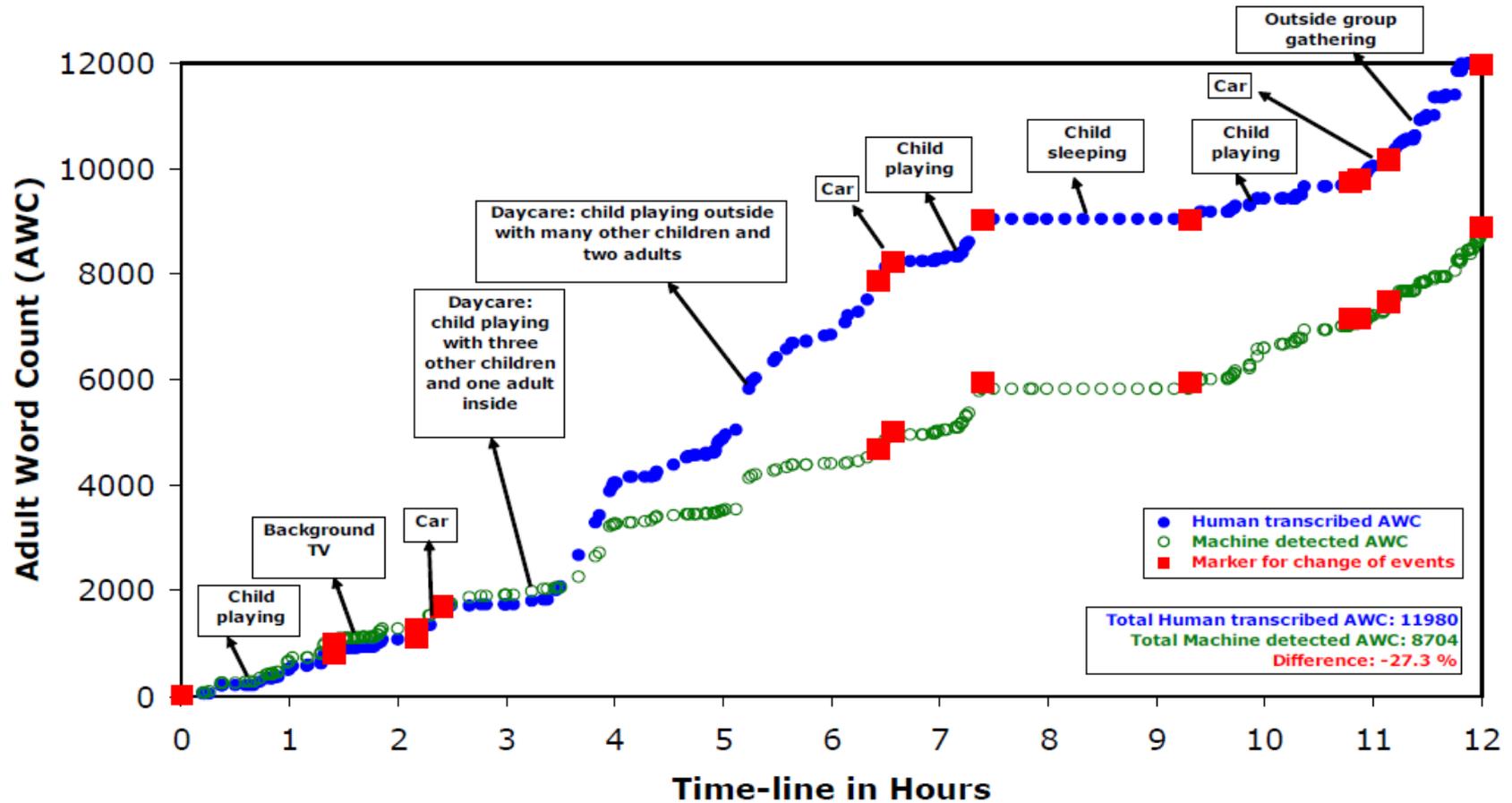- Child Voc
  - r = .71 (Canault et al. 2015, French)
  - LENA:

	Child Voc	Veg/Cry
Child Voc	75%	25%
Veg/Cry	16%	84%

LENA's segment labeling

	Adult	Child	TV	Other
Adult	82%	2%	4%	12%
Child	7%	76%	0%	17%
TV	8%	0%	71%	21%
Other	14%	4%	6%	76%

LENA under noisy conditions (LTR -5, Fig 3)

How does LENA fare in noisy settings?

- Canault et al.: AWC impacted

- Soderstrom & Wittebolle (2013)
  - AWC
    - Daycare r = .77
    - Home r = .83
  - Child Voc
    - Daycare r = .72
    - Home r = .65

General Questions

- What is the impact of noisy environments on LENA's quality?

- How good is LENA's *labeling* of segments? What are its strengths and weaknesses?
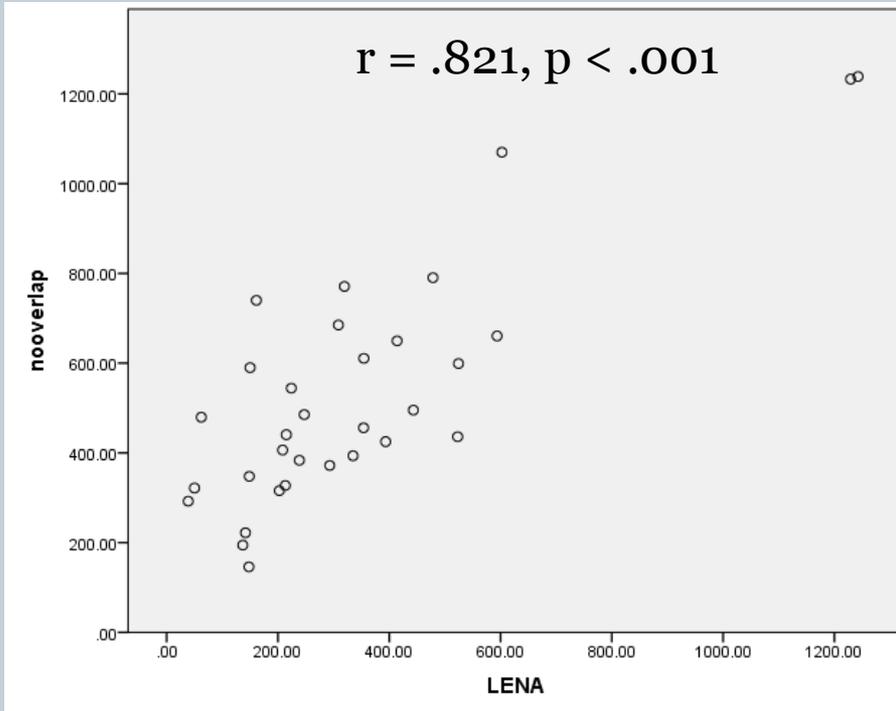
Participants

	N participants	Ages (months)	N files	LENA segs
Home	12	12-30	30	7527
Home Daycare	7	14-30	21	4944
Daycare Centre	13	14-38	36	7101

Transcripts

- 15 minute sections transcribed 1 per day
- 1 hr into recording (mid-morning)
- Each file
  - Transcribed separately by two transcribers in Transcriber
  - Consensus final transcript
  - Start from LENA segmentation
    - Codes to indicate if disagreement with LENA (on segmentation)
    - By word
  - LENA labeling present but not used by transcribers
  - Coding: Speaker, Addressee, Utterance characteristics

AWC: Overall

r = .821, p < .001

r = .757, p < .001

AWC: By ChildCare Setting

r = .948, p < .001

r = .677, p = .095

r = .855, p < .001

No Overlap

r = .929, p < .001

r = .691, p = .087

r = .722, p = .005
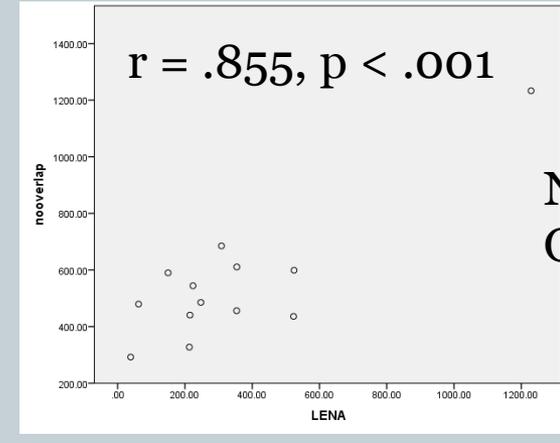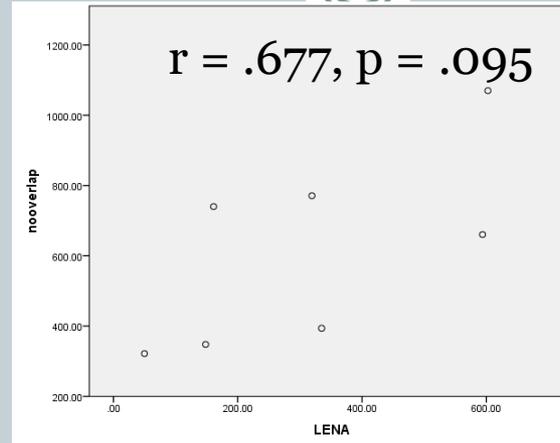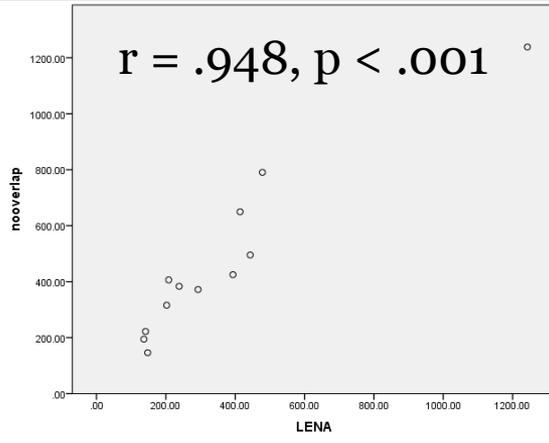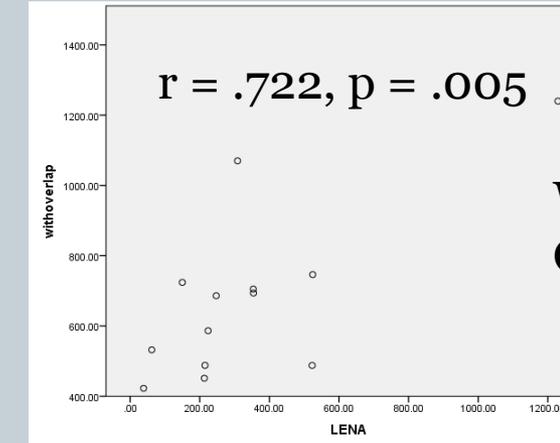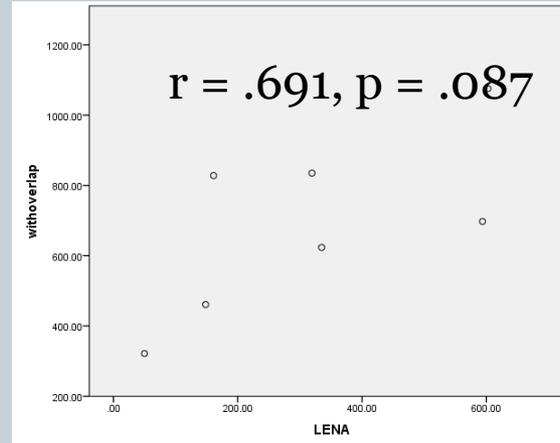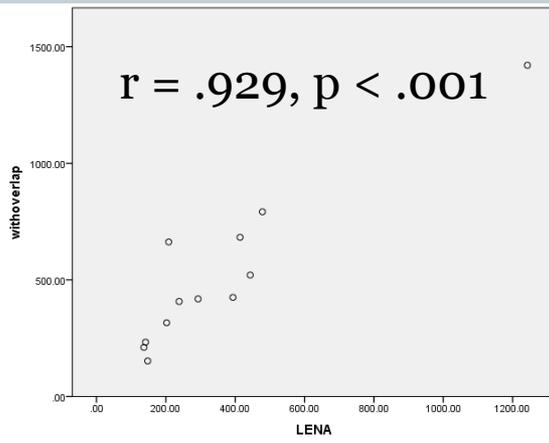
With Overlap

Home   Home Daycare   Daycare Centre

Number of adult words / 5 min (LENA: ⬤, transcribed: ⬤)
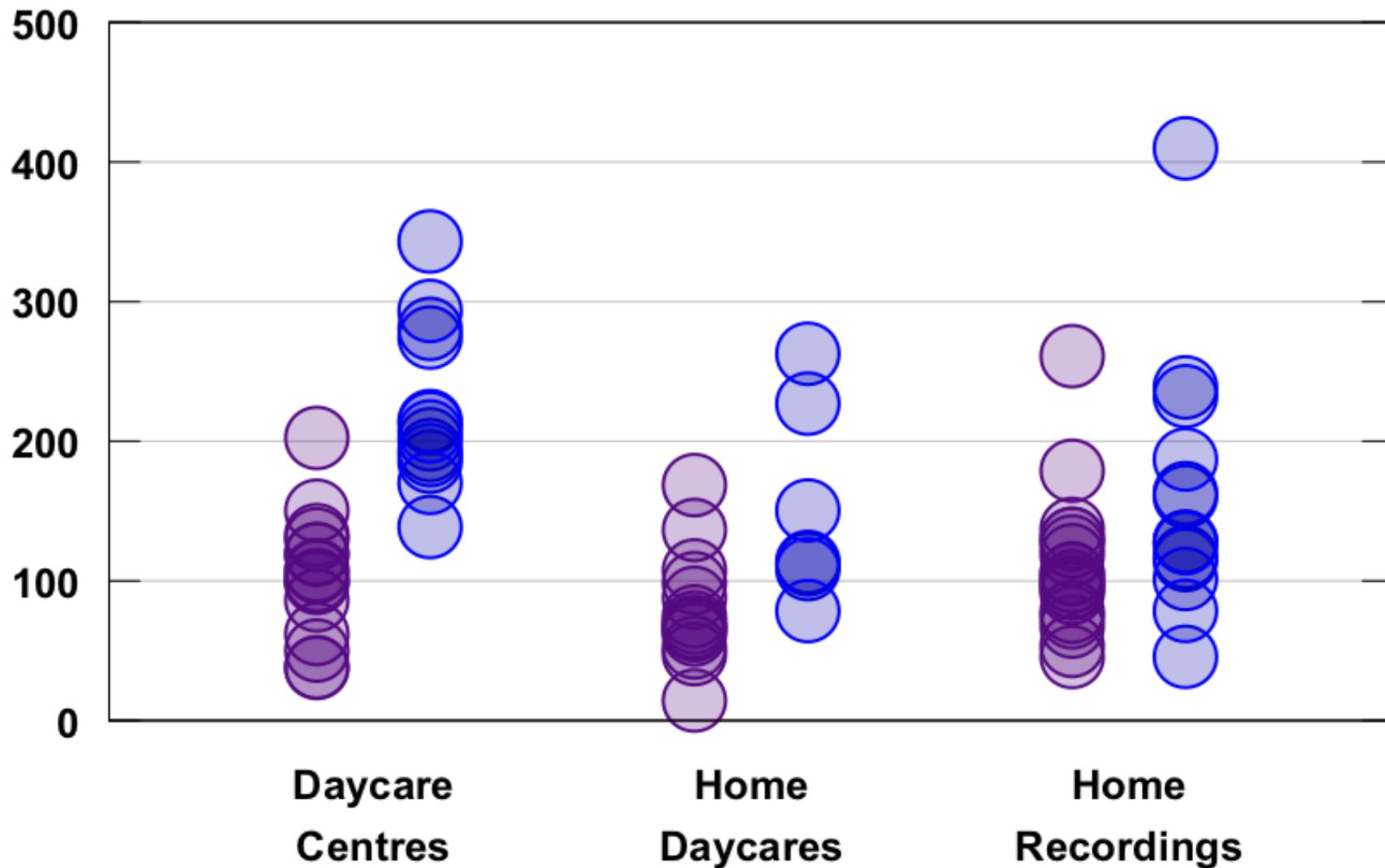
Labeling Confusion Matrices

LENA	Transcriber
Target child	Target child
Other child	Other child
	Child-uncertain
Female adult	Female adult
Male adult	Male adult
	Uncertain
Overlap (speech + speech/ speech + noise)	Overlap (speech + speech only)
Non-speech	Non-speech > 1s
	Non-speech < 1s

Row-wise percent scores: Overall

	Target Child	Other Child	Female Adult	Male Adult	Overlapping speech	Non-speech
Target Child	55 (2860)	11 (582)	3 (153)	0 (9)	20 (1046)	11 (564)
Other Child	9 (261)	46 (1340)	4 (112)	0 (0)	34 (999)	7 (212)
Child (Undefined)	34 (485)	22 (316)	2 (34)	0 (1)	32 (454)	10 (138)
Female Adult	5 (467)	8 (874)	36 (3705)	2 (156)	40 (4193)	9 (980)
Male Adult	4 (34)	1 (8)	9 (69)	36 (277)	42 (324)	8 (64)
Uncertain	8 (2)	32 (8)	4 (1)	0 (0)	48 (12)	8 (2)
Overlapping Speech	12 (170)	9 (131)	15 (209)	0 (6)	63 (894)	1 (20)
non-speech (>=1sec)	2 (259)	3 (490)	2 (301)	1 (157)	31 (4580)	61 (9079)
non-speech (<1sec)	9 (1564)	6 (1011)	1 (173)	0 (11)	21 (3868)	64 (11546)

0-.25
.26-.49
.5-.74
.75-1.0

Column-wise percent scores: Overall

	Target Child	Other Child	Female Adult	Male Adult	Overlapping speech	Non-speech
Target Child	47 (2860)	12 (582)	3 (153)	1 (9)	6 (1046)	2 (564)
Other Child	4 (261)	28 (1340)	2 (112)	0 (0)	6 (999)	1 (212)
Child (Undefined)	8 (485)	7 (316)	1 (34)	0 (1)	3 (454)	1 (138)
Female Adult	8 (467)	18 (874)	78 (3705)	25 (156)	26 (4193)	4 (980)
Male Adult	1 (34)	0 (8)	1 (69)	45 (277)	2 (324)	0 (64)
Uncertain	0 (2)	0 (8)	0 (1)	0 (0)	0 (12)	0 (2)
Overlapping Speech	3 (170)	3 (131)	4 (209)	1 (6)	5 (894)	0 (20)
non-speech (>=1sec)	4 (259)	10 (490)	6 (301)	25 (157)	28 (4580)	40 (9079)
non-speech (<1sec)	26 (1564)	21 (1011)	4 (173)	2 (11)	24 (3868)	51 (11546)

0-.25
.26-.49
.5-.74
.75-1.0

Overlap

- Major influencing difference transcribers and LENA
  - Row-wise 20-48% coded by LENA as overlap
- Comparing across settings (row-wise)

	Home	HD	DC
Female adult	28%	35%	49%
Target child	16%	17%	27%

- Column-wise: LENA's overlap judged by transcribers

	Home	HD	DC
Overlap	3%	3%	8%
Non-speech	62%	56%	44%

FAN-Female Adults

Row-wise	Home	HD	DC
Total percent	43%	37%	31%
Excluding overlap	60%	57%	61%

Column-wise	Home	HD	DC
Total percent	76%	71%	83%

CHN-Target Child

Row-wise	Home	HD	DC
Total percent	58%	57%	50%
Excluding overlap	68%	69%	69%

Column-wise	Home	HD	DC
Target Child	57%	43%	40%
Other Child	0%	14%	3%
Child Uncertain	5%	4%	13%
Female Adult	7%	3%	11%
< 1 s non-speech	24%	31%	24%

Male Adult

Row-wise	Home	HD	DC
Total percent	38%	0%*	29%
Excluding overlap	62%	0%*	62%

Column-wise	Home	HD	DC
Male Adult	54%	0%*	28%
Female Adult	17%	58%*	40%
Non-speech	26%	38%*	29%

What does this mean?

- ## What does it say about LENA?
  - Best for female adult
  - Errs on the side of exclusion, oversegmenting
  - Swamping of "good" tokens in low-frequency types
- ## Comparing machine and human
  - Humans are really good – maybe too good?
  - Identification of target vs. other child, male vs. female
- ## What about home (quiet) vs. noise (daycare)?
  - Not as different as might be expected
  - Expected effects on AWC, overlapping speech

Acknowlegments

- Many many RAs for data collections/transcription
- Parents, teachers, daycares, children who particiapted

Row-wise percent scores: Home

	Target Child	Other Child	Female Adult	Male Adult	Overlapping speech	Non-speech
Target Child	58 (1302)	12 (279)	4 (92)	0 (6)	16 (351)	10 (222)
Other Child	4 (5)	12 (14)	7 (8)	0 (0)	57 (65)	20 (23)
Child (Undefined)	63 (119)	11 (21)	1 (2)	0 (0)	12 (22)	14 (26)
Female Adult	5 (156)	7 (211)	43 (1339)	2 (74)	28 (889)	14 (452)
Male Adult	5 (31)	1 (7)	10 (63)	38 (232)	39 (240)	7 (44)
Uncertain	0 (0)	14 (1)	0 (0)	0 (0)	86 (6)	0 (0)
Overlapping Speech	12 (31)	9 (23)	20 (52)	2 (5)	54 (142)	4 (10)
non-speech (>=1sec)	1 (91)	1 (99)	2 (146)	2 (103)	22 (1480)	71 (4795)
non-speech (<1sec)	8 (540)	3 (177)	1 (62)	0 (7)	19 (1305)	69 (4731)

Legend:
- 0-.25
- .26-.49
- .5-.74
- .75-1.0

Row-wise percent scores: Home Daycare

	Target Child	Other Child	Female Adult	Male Adult	Overlapping speech	Non-speech
Target Child	57 (574)	13 (135)	1 (14)	0 (1)	17 (172)	11 (111)
Other Child	9 (184)	53 (1100)	4 (94)	0 (0)	27 (571)	7 (144)
Child (Undefined)	14 (47)	41 (141)	3 (10)	0 (0)	36 (126)	6 (22)
Female Adult	2 (34)	14 (260)	37 (691)	1 (18)	35 (656)	11 (211)
Male Adult	0 (0)	0 (0)	0 (0)	0 (0)	50 (2)	50 (2)
Uncertain	7 (1)	50 (7)	7 (1)	0 (0)	21 (3)	14 (2)
Overlapping Speech	7 (18)	25 (69)	21 (56)	0 (0)	45 (124)	2 (6)
non-speech (>=1sec)	2 (61)	7 (244)	1 (54)	0 (10)	29 (1062)	61 (2268)
non-speech (<1sec)	9 (422)	9 (416)	1 (47)	0 (2)	22 (1051)	60 (2916)

0-.25
.26-.49
.5-.74
.75-1.0

Row-wise percent scores: Daycare Centre

	Target Child	Other Child	Female Adult	Male Adult	Overlapping speech	Non-speech
Target Child	50 (984)	9 (168)	2 (47)	0 (2)	27 (523)	12 (231)
Other Child	10 (72)	32 (226)	1 (10)	0 (0)	51 (363)	6 (45)
Child (Undefined)	36 (319)	17 (154)	2 (22)	0 (1)	34 (306)	10 (90)
Female Adult	5 (277)	7 (403)	31 (1675)	1 (64)	49 (2648)	6 (317)
Male Adult	2 (3)	1 (1)	4 (6)	29 (45)	53 (82)	12 (18)
Uncertain	25 (1)	0 (0)	0 (0)	0 (0)	75 (3)	0 (0)
Overlapping Speech	14 (121)	4 (39)	11 (101)	0 (1)	70 (628)	0 (4)
non-speech (>=1sec)	2 (107)	3 (147)	2 (101)	1 (44)	46 (2038)	45 (2016)
non-speech (<1sec)	9 (602)	6 (418)	1 (64)	0 (2)	23 (1512)	60 (3899)

0-.25
.26-.49
.5-.74
.75-1.0

Column-wise percent scores: Home

	Target Child	Other Child	Female Adult	Male Adult	Overlapping speech	Non-speech
Target Child	57 (1302)	34 (279)	5 (92)	1 (6)	8 (351)	2 (222)
Other Child	0 (5)	2 (14)	0 (8)	0 (0)	1 (65)	0 (23)
Child (Undefined)	5 (119)	3 (21)	0 (2)	0 (0)	0 (22)	0 (26)
Female Adult	7 (156)	25 (211)	76 (1339)	17 (74)	20 (889)	4 (452)
Male Adult	1 (31)	1 (7)	4 (63)	54 (232)	5 (240)	0 (44)
Uncertain	0 (0)	0 (1)	0 (0)	0 (0)	0 (6)	0 (0)
Overlapping Speech	1 (31)	3 (23)	3 (52)	1 (5)	3 (142)	0 (10)
non-speech (>=1sec)	4 (91)	12 (99)	8 (146)	24 (103)	33 (1480)	47 (4795)
non-speech (<1sec)	24 (540)	21 (177)	4 (62)	2 (7)	29 (1305)	46 (4731)

0-.24
.25-.49
.5-.74
.75-1.0

Column-wise percent scores: Home Daycare

	Target Child	Other Child	Female Adult	Male Adult	Overlapping speech	Non-speech
Target Child	43 (574)	6 (135)	1 (14)	3 (1)	5 (172)	2 (111)
Other Child	14 (184)	46 (1100)	10 (94)	0 (0)	15 (571)	3 (144)
Child (Undefined)	4 (47)	6 (141)	1 (10)	0 (0)	3 (126)	0 (22)
Female Adult	3 (34)	11 (260)	71 (691)	58 (18)	17 (656)	4 (211)
Male Adult	0 (0)	0 (0)	0 (0)	0 (0)	0 (2)	0 (2)
Uncertain	0 (1)	0 (7)	0 (1)	0 (0)	0 (3)	0 (2)
Overlapping Speech	1 (18)	3 (69)	6 (56)	0 (0)	3 (124)	0 (6)
non-speech (>=1sec)	5 (61)	10 (244)	6 (54)	32 (10)	28 (1062)	40 (2268)
non-speech (<1sec)	31 (422)	18 (416)	5 (47)	6 (2)	28 (1051)	51 (2916)

Legend
0-.25
.26-.49
.5-.74
.75-1.0

Column-wise percent scores: Daycare Centre

	Target Child	Other Child	Female Adult	Male Adult	Overlapping speech	Non-speech
Target Child	40 (984)	11 (168)	2 (47)	1 (2)	6 (523)	3 (231)
Other Child	3 (72)	15 (226)	0 (10)	0 (0)	4 (363)	1 (45)
Child (Undefined)	13 (319)	10 (154)	1 (22)	1 (1)	4 (306)	1 (90)
Female Adult	11 (277)	26 (403)	83 (1675)	40 (64)	33 (2648)	5 (317)
Male Adult	0 (3)	0 (1)	0 (6)	28 (45)	1 (82)	0 (18)
Uncertain	0 (1)	0 (0)	0 (0)	0 (0)	0 (3)	0 (0)
Overlapping Speech	5 (121)	3 (39)	5 (101)	1 (1)	8 (628)	0 (4)
non-speech (>=1sec)	4 (107)	9 (147)	5 (101)	28 (44)	25 (2038)	30 (2016)
non-speech (<1sec)	24 (602)	27 (418)	3 (64)	1 (2)	19 (1512)	59 (3899)

0-.25
.26-.49
.5-.74
.75-1.0