

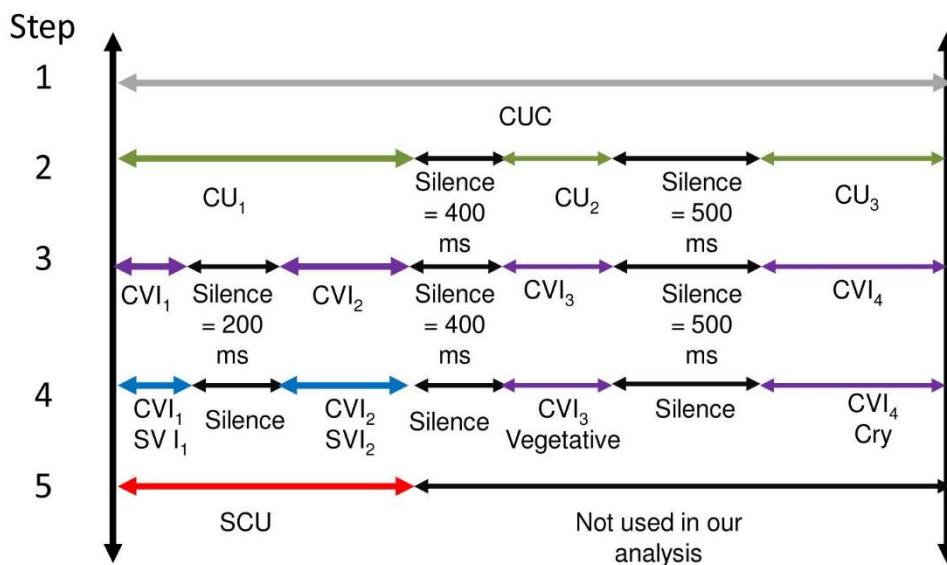
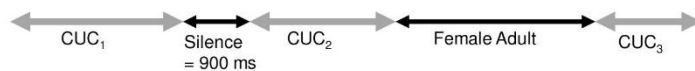
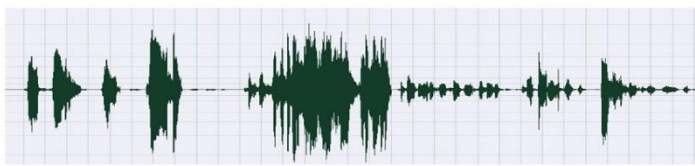
期待 77-喃語と ASD:

Oller, et al. (2010) PNAS, 107:13354-13359 と Supporting Information Appendix

この論文は、音声を自動的に分析する LENA システムを利用して定型発達、言語遅滞、ASD の音声発達を比較した。音声の分析手続きを簡単に述べる。幼児の音声を CUC: Child Utterance Cluster に分けるが、800 ms 以上の無音区間で区切る。CUC は 600 ms 以上の持続を持つ。大人の声や他の音が入ると、除外する (上図)。次に、CUC を CU に分ける。CU は 300 ms 以上で、300 ms の無音区間で区切られる。CU は CVI: Speech Related Vocal Island に分けられる。CVI は少なくとも 50 ms の rise-fall で 300 ms 以内の持続を持つ。CVI の内

Vegetative と Cry を除いた CVI が SVI: Speech Related Vocal Island で SVI を含む CU が SCU: Speech Related Child Utterance で分析の対象となる。Vegetative、Cry は分析しない (下図)。SVI を含む CU 内に Cry 等があれば、それも分析

から除外する。

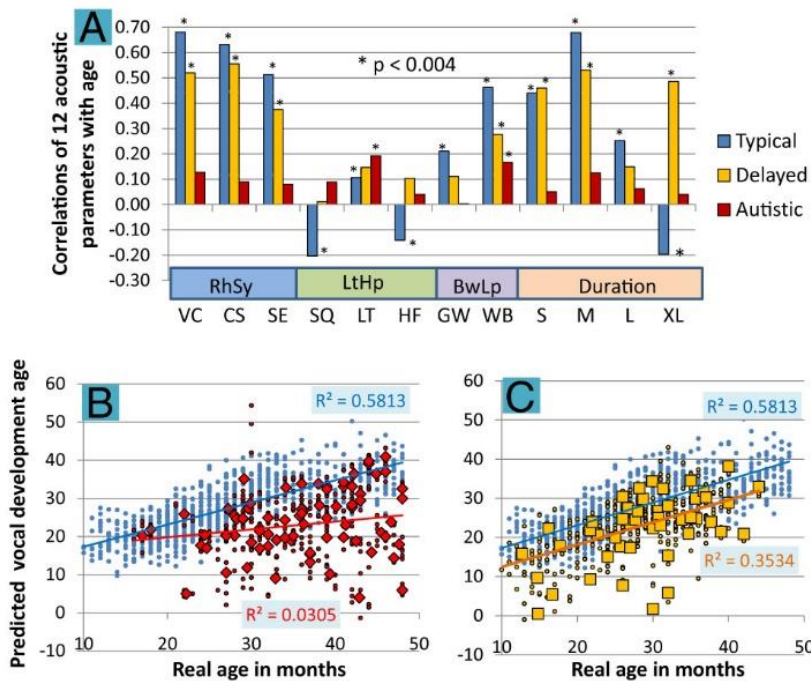


LENA が SVI の分析に利用した 12 の音声パラメータが下の表である。説明は省略する。12 のパラメータに関して、有 yes/無 no、+/- のラベルを貼った。

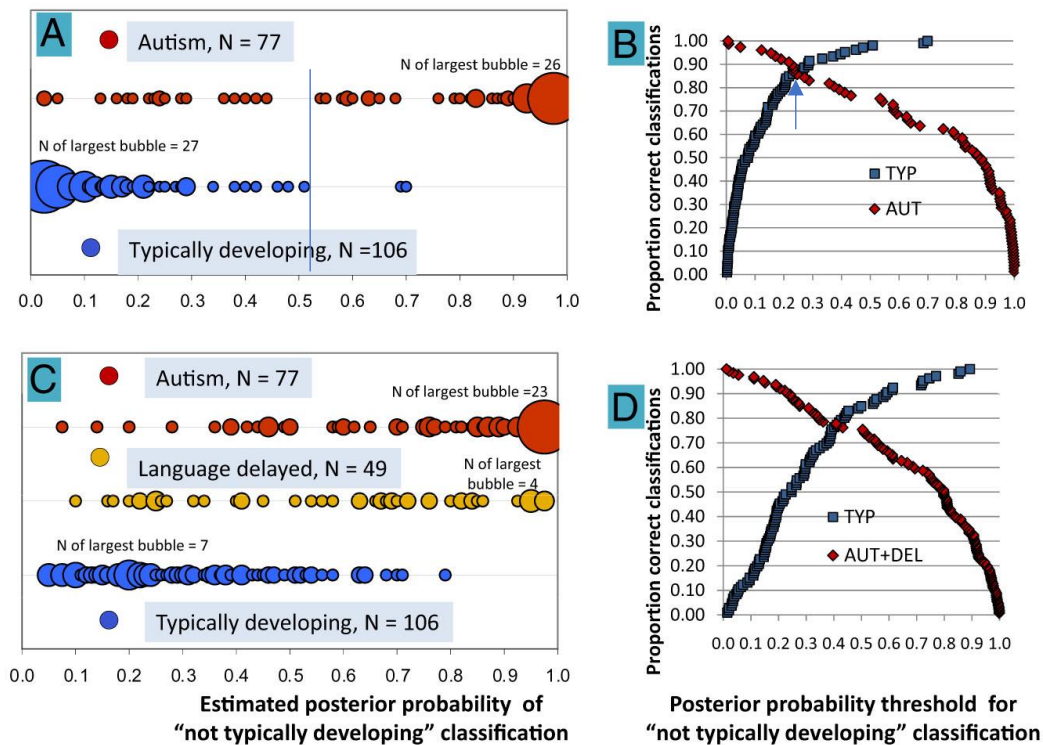
a. Rhythm/Syllabicity			Positive classification on group a features suggested speech-like rhythmic organization because values analyzed were islands (roughly, syllables) per utterance (SVIs per SCU) showing group a features. Thus, utterances were rhythmically organized in accord with speech if they tended to show relatively high numbers of syllables per utterance with voicing, canonical formant transitions, and spectral entropy variations typical of speech.
1	VC	Voiced or unvoiced: pitch detectable through > 50% of SVI (roughly, syllable)	
2	CS	Canonical Syllable transitions or not: Formant (F1 and F2) transitions < 120 ms	
3	SE	Spectral Entropy typical of speech or not	
b. Low spectral tilt and high pitch control			Positive classification on group b parameters suggested control of high pitch and low spectral tilt, which tend to occur in certain typical emotional expressions of high intensity (squeal quality). More islands per utterance positive on b parameters suggested more active emotional expression in the high spectral frequency range.
4	SQ	Mean pitch high or not (SQueal): > 600 Hz	
5	LT	Low Tilt of spectrum or not	
6	HF	High Frequency energy concentration or not	
c. Wide formant bandwidth and low pitch control			Positive classification on group c parameters suggested control of low pitch and high bandwidths of the first two formants, qualities which tend to occur in certain typical emotional expressions of high intensity (growl quality). More islands per utterance positive on c parameters suggested more active emotional expression in the low spectral frequency range.
7	GW	Mean pitch low or not (GroWl): < 250 Hz	
8	WB	Wide Bandwidth of first two formants or not	
d. Duration of islands (SVIs) within utterances (SCUs)			Group d parameters split according to durations typical of syllables in speech. Positive classification on parameters 9 and 10 suggested speech-like rhythmic organization because the durational values indicated are typical of syllables in speech. More islands per utterance with 9 and 10 thus suggested more speech-like syllables. Positive classification on parameters 11 and 12 suggested the opposite, because the corresponding ranges are beyond the durations of typical syllables.
9	S	Short (110 - 250 ms)	
10	M	Medium (250 - 600 ms)	
11	L	Long (600 - 900 ms)	
12	XL	EXtra Long (900 - 3000 ms)	

録音時間や、発声数が幼児によって異なるので、+のついた SVI の数と SCU の数の比 SVI/SCU ratio をパラメータごとに求めた。

次ページの上図 A は 12 のパラメータの SVI/SCU と年齢の相関を定型発達 Typical, 言語遅滞 Delayed, 自閉スペクトラム障がい Autistic に分けて表示した。4 つのカテゴリ内では比較的類似の結果になっている。また、例外はあるものの、Autistic では全般的に相関が低い。図 B, C は Typical の音声の発達的变化と、Autistic 赤と Delayed 黄の比較をした重回帰分析の結果である。図 C の Delayed では回帰直線は Typical の下だが平行の関係にあり、同じように発達しているが、一方、図 B の Autistic では年齢を重ねるに従い変動が大きくなり、Typical の発達のラインから離れていくのが分かる。

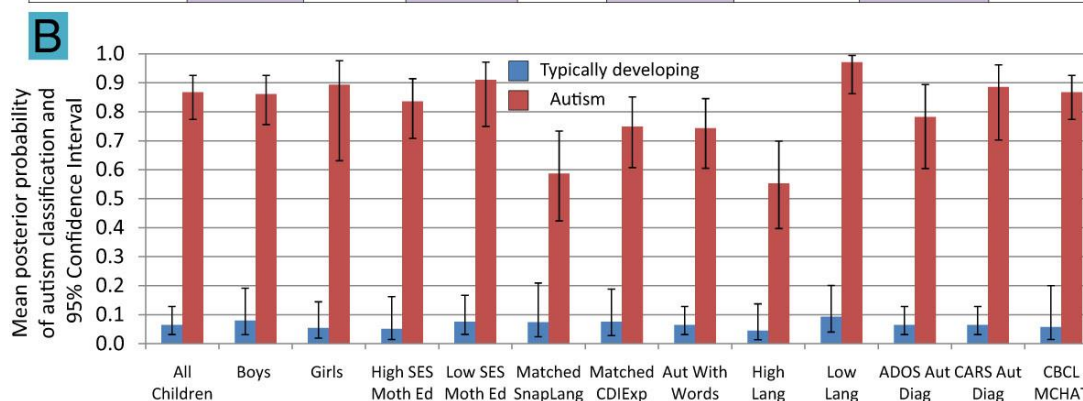


その後は、音声データで3つの幼児の群を分類できるかを、一個抜き交叉検証 LOOCV による線形判別分析 LDA で検討した。下図はその結果で、図 A, B は Autistic と Typical の分類で、横軸は「Typical でない」の推定事後確率。事後確率の cutoff (閾値) が 0.52 で、sensitivity (hit rate) は 0.75, specificity (correct rejection rate) は 0.98, 等エラー確率 (図 B の交点) での sensitivity と specificity は 0.86 だった。下図 C, D は Delayed を加えた結果



である。0.56 の cutoff で、Typical が正しく同定されたのは 0.90, Autistic と Delayed が Typical と分類されなかったのはそれぞれ 0.80, 0.62 だった。等エラー確率 (図 D の交点) での sensitivity と specificity は 0.79 だった。下の表は 4 種類のデータで 3 群の幼児のすべての組み合わせで分類を行った結果である。上の説明は左端の Entire sample の結果である。いずれのデータでも多くの組み合わせで分類は有意で、LENA system が有効であることを示した。以下省略する。

A	Entire sample, N=232, LOOCV		Matched sample, N=113, LOOCV		Train on Phase I data (N=138), test on Phase II data (N=94)		Train on all recordings, test on first recording, N=232, LOOCV		
	Configuration	Sens/Spec	$\chi^2, p<$	Sens/Spec	$\chi^2, p<$	Sens/Spec	$\chi^2, p<$	Sens/Spec	$\chi^2, p<$
	Aut vs Typ	0.86	5.5E-22	0.82	2.4E-08	0.87	5.0E-10	0.83	6.7E-19
	Aut vs Typ + Del	0.79	8.7E-18	0.74	1.1E-06	0.73	1.3E-05	0.78	8.0E-16
	Typ vs Aut + Del	0.79	5.4E-19	0.76	2.4E-07	0.80	2.8E-08	0.77	9.0E-17
	Typ vs Del	0.73	1.1E-07	0.69	6.8E-04	0.73	0.001	0.68	3.4E-05
	Aut vs Del	0.70	9.7E-06	0.66	0.007	0.67	0.0117	0.66	5.0E-04
	Del vs Typ + Aut	0.60	0.017	0.54	0.424	0.57	0.247	0.50	0.936



結局、LENA を導入することは、音声の分析に必要な人員と時間を節約することになるかもしれない。ただ、分析は大まかな可能性があり、研究の目的により音声分析専門家のチェックが必要になるだろう。また、ビデオ撮影による視覚データはどうしても欠かせない。LENA には専用のマイクがあるらしいが、LENA がビデオの音声データを分析できるか分からない。いずれにしてもビデオと併用することが好ましい。

LENA を利用した研究は Google scholar で調べることができるだろう。この論文と同じ時期の Oller のグループの研究に、Warren et al. (2010) J. Autism Dev. Disord., 40:555-569 がある。また、新しい論文では、Swanson et al. (2019) Autism Res., 12:1784-1795 がある (この論文は来月に紹介する予定である)。LENA の会社の HP もあるようだ。