最先端NLP9

Probabilistic Typology: Deep Generative Models of Vowel Inventories

Ryan Cotterell and **Jason Eisner**
Department of Computer Science
Johns Hopkins University
{ryan.cotterell,eisner}@jhu.edu

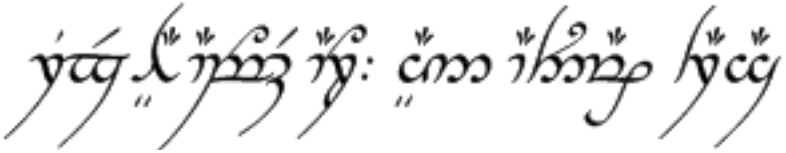
ACL 2017

持橋大地
統計数理研究所
daichi@ism.ac.jp

2017-9-16 (土)

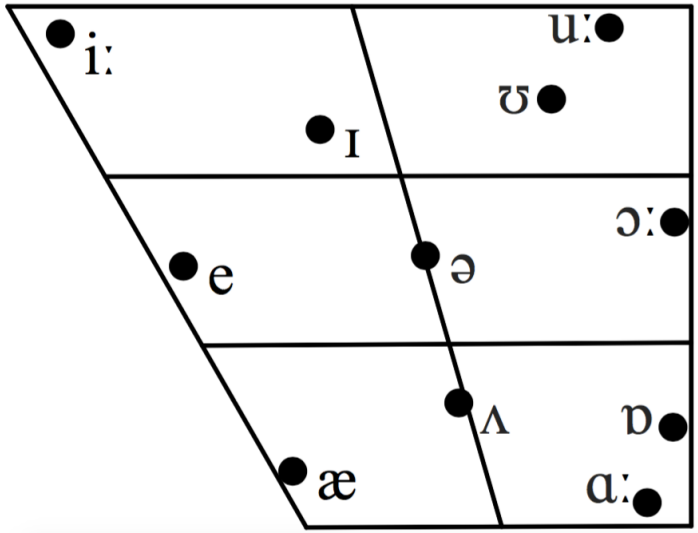
※ 図はすべて、著者のプレゼンテーションから

背景と動機

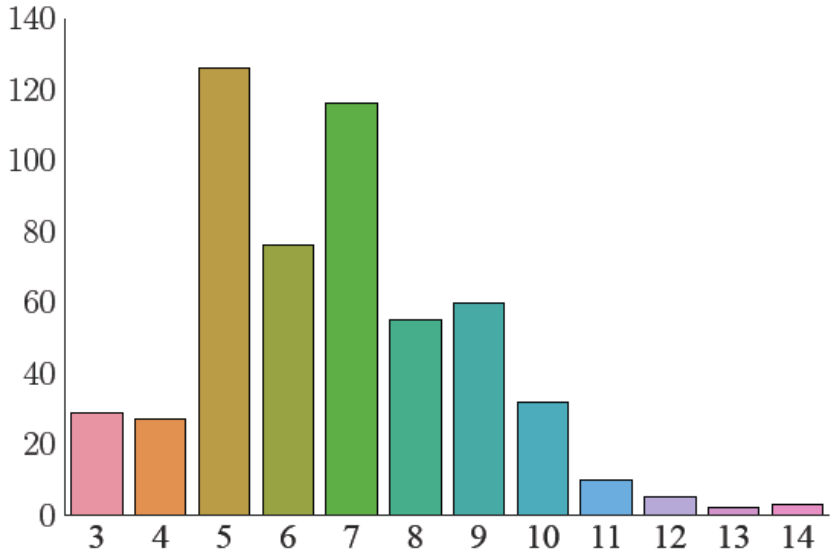
- 類型論 (Typology): 様々な言語に共通する性質を探る
- 理想的には、「新しい言語を生成できる」確率分布を求めたい
  - Qwenya (「指輪物語」) 
  - クリンゴン (Star Trek)　bI'avtaHvIS jot'a'?
  - Dothraki (Game of Thrones) Jadi anna, zhey yalli anni.

この論文

- 各言語の持つ母音の集合をモデル化



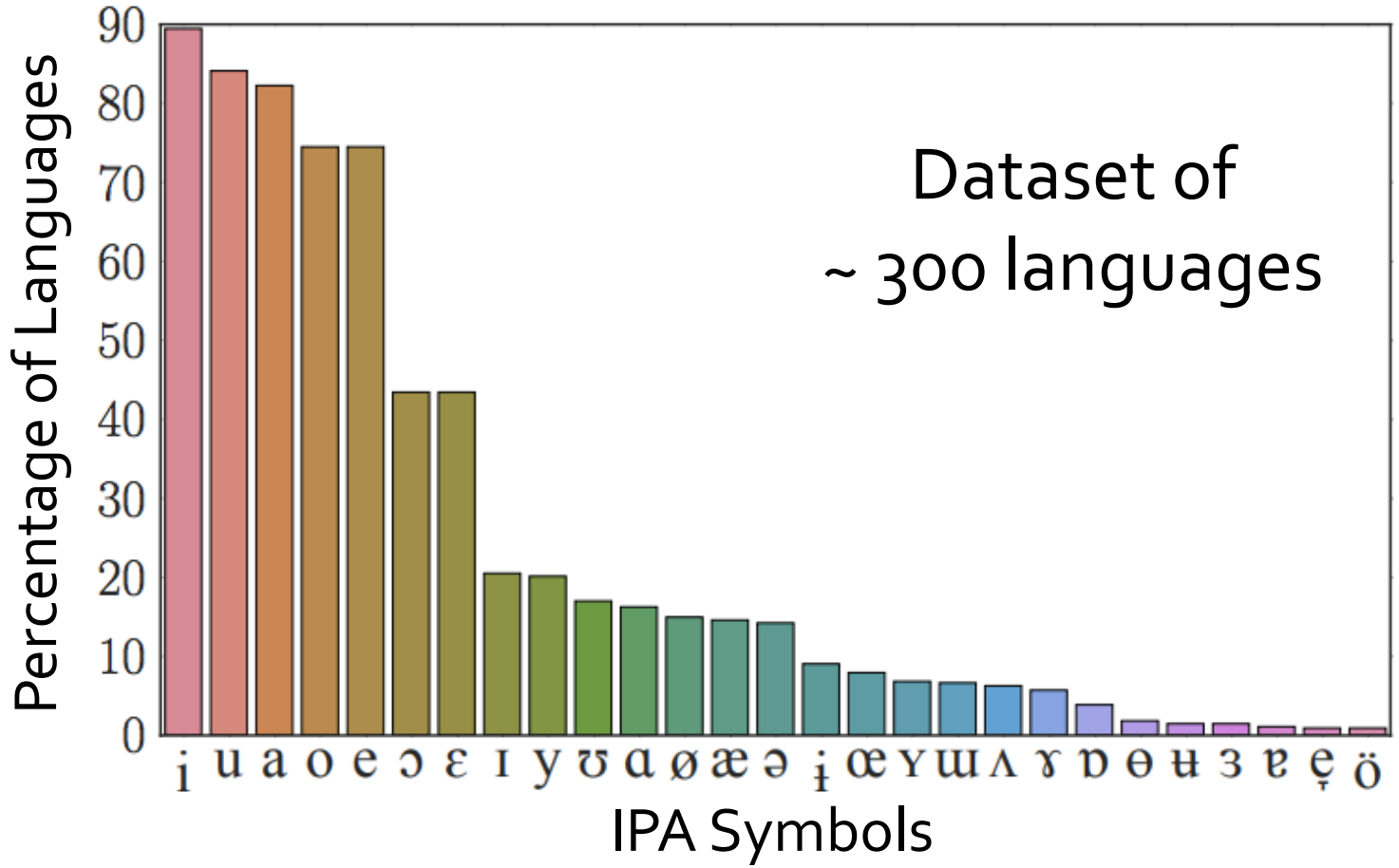
conventional drawing



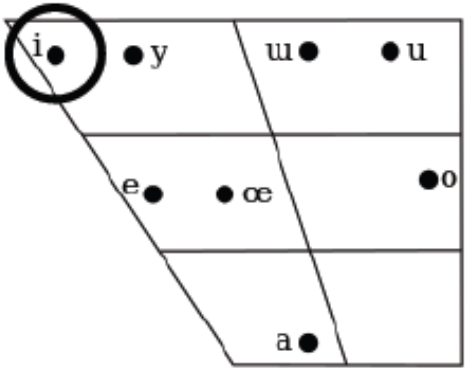
- 母音数の実際の分布

言語と母音

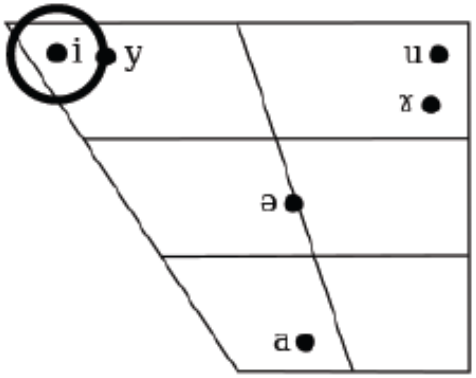
- 実験データでのIPA母音と、それを持つ言語の数



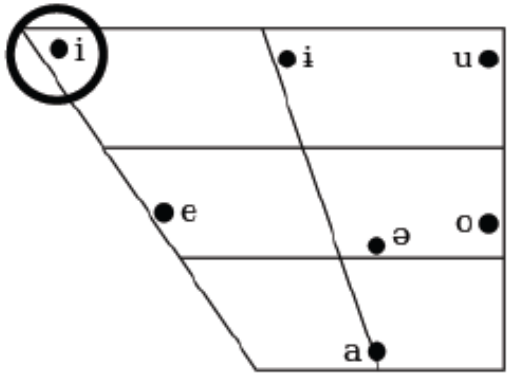
Example Vowel Systems



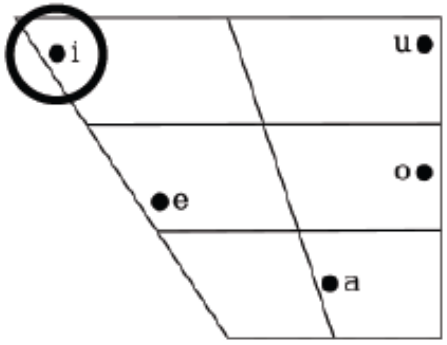
Turkish



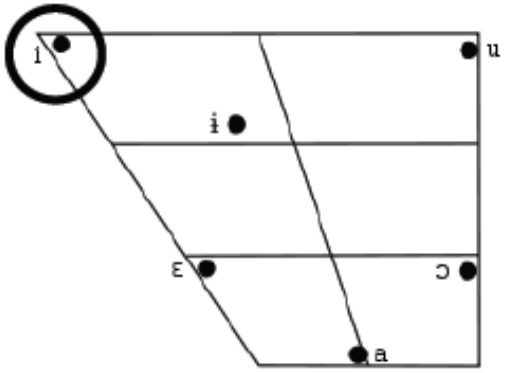
Chinese



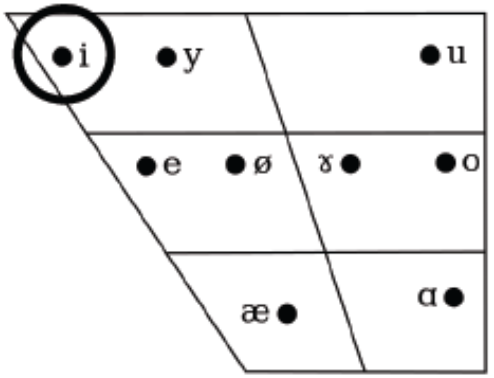
Romanian



Quechua



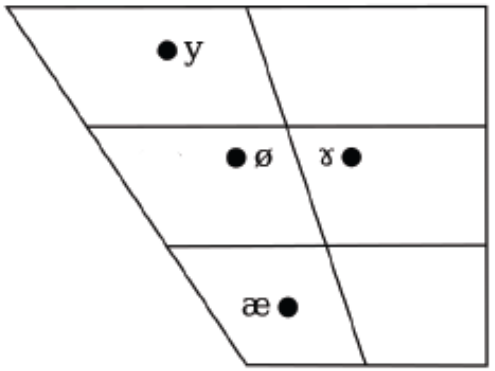
Polish



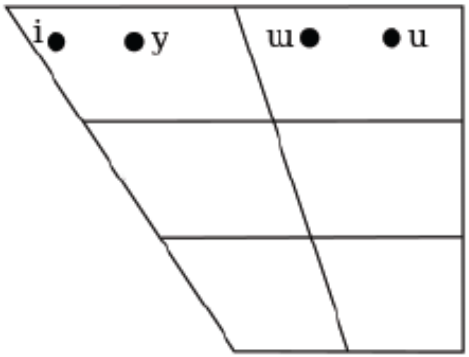
Bulgarian

Unattested Vowel Systems

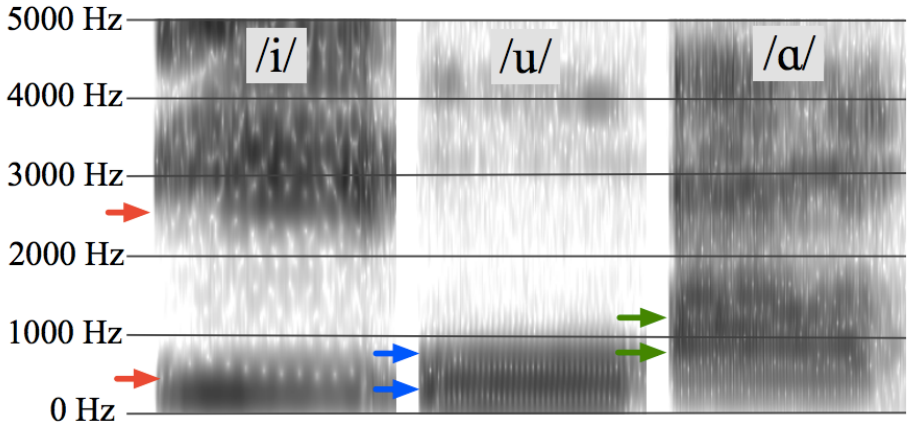
Mostly Infrequent Vowels



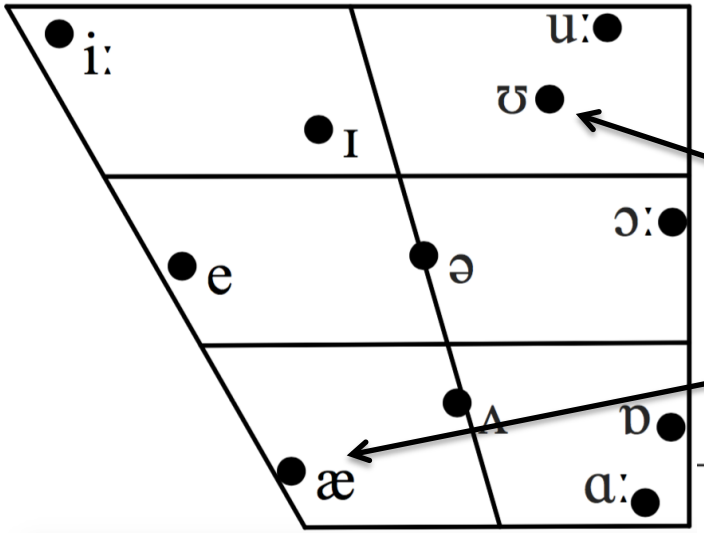
Vowels not Spread Out



物理的には:



- 音声をフーリエ変換した際の係数 $F_0, F_1, F_2, \cdots$ について、F1,F2 を特徴として採用
  - F0は波の基本周波数で、共通



conventional drawing

$$\boldsymbol{f}(\upsilon) = [450, 1030]$$

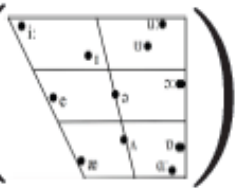
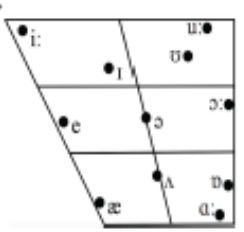
$$\boldsymbol{f}(\text{æ}) = [690, 1660]$$

母音の分布 (言語学)

- Dispersion (分散)
  母音同士は、他の母音と弁別できるように離れているべき
- Focalization (焦点化)
  発音しやすいよう、F1とF2が近い母音が選ばれやすい

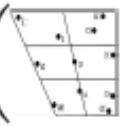
Dispersion-Focalization Theory (Schwartz+ (1997))
上の2つの条件を両方満たす言語が望ましい

The Generative Model

$$p(\text{  }) \propto \text{score}\left(\text{  }\right)$$

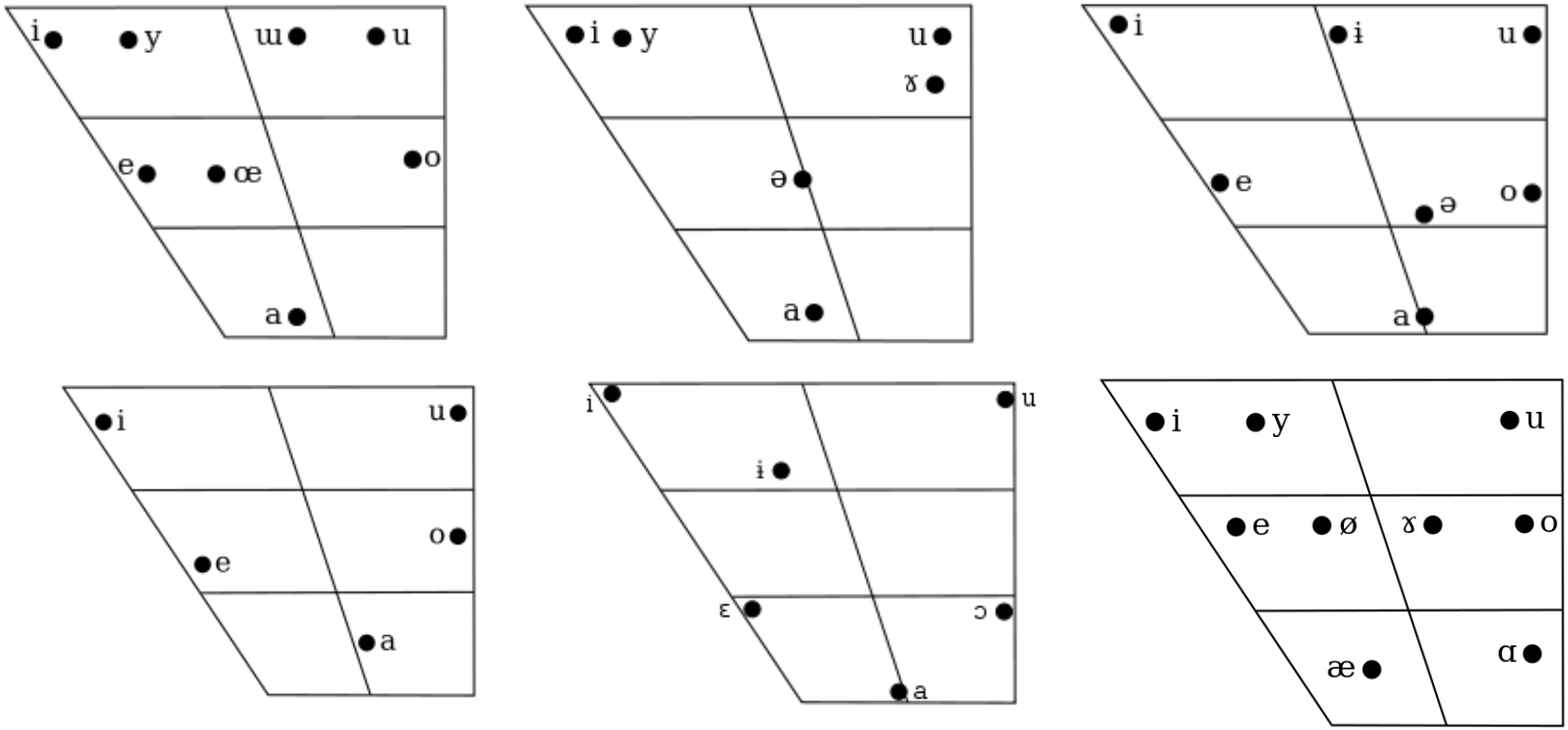
probability measure

"goodness" of a vowel system

- How do we define $\text{score}\left(\text{  }\right)$?
  - How do we quantify focalization and dispersion?
- Question in linguistics, but we use **tools from NLP/ML**

母音の分布 (2)

- 数学的には、ポアソン点過程
  - 雨、地震、疫病、商品の購買・・・・



母音の分布 (3)

- 3つのモデル化の方法
  (1) ベルヌーイ点過程
  (2) マルコフ確率場
  (3) 行列式点過程

(1) ベルヌーイ点過程 (BPP)

$$p(V) \propto \prod_{v \in V} \psi(v)$$

- 母音を、重要度(ポテンシャル) $\phi(v)$ に応じて独立に選ぶ
- 最も簡単なベースライン
  - 母音どうしが離れているという制約がない

(2) マルコフ点過程 (MPP)

$$p(V) \propto \prod_{v \in V} \psi(v) \prod_{v,w \in V} \psi(v,w)$$

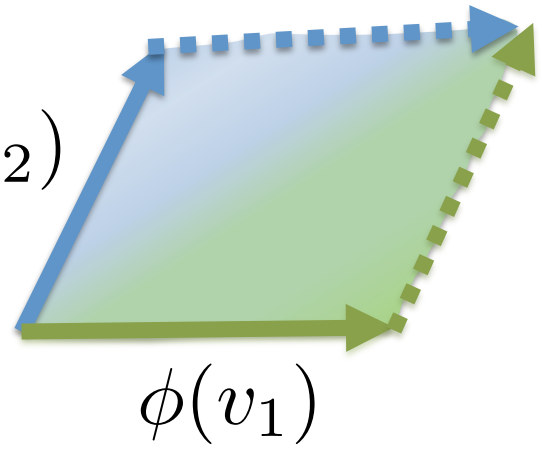
- 機械学習の言葉では、マルコフ確率場(MRF)
  - ただし、全連結なことに注意
- BPPに加え、母音vとwの間にポテンシャル $\psi(v,w)$ が存在
  - “Repulsion”を表現できる

(3) 行列式点過程 (DPP)

- Determinantal Point Process
  - 機械学習に最近導入 (NIPS 2010あたり)
- 行列式の値＝特徴ベクトルの張る平行n面体の体積^2に点集合の確率が比例

$$p(V) \propto \det L_V$$

$$L_V = (\phi(v_1), \cdots, \phi(v_M))$$



  - できるだけ直交している点集合が選ばれる
  - 正規化定数は求まる: $\det(L + I)$

行列式点過程 (2)

$$p(V) \propto \det L_V$$

$$L_V = (\phi(v_1), \cdots, \phi(v_M))$$

- $\phi(v)$ のノルムは制限していないので、特定の母音が重要なことを表現できる

↓

- Dispersion-Focalizationの両方を表現できる！

Deep point process

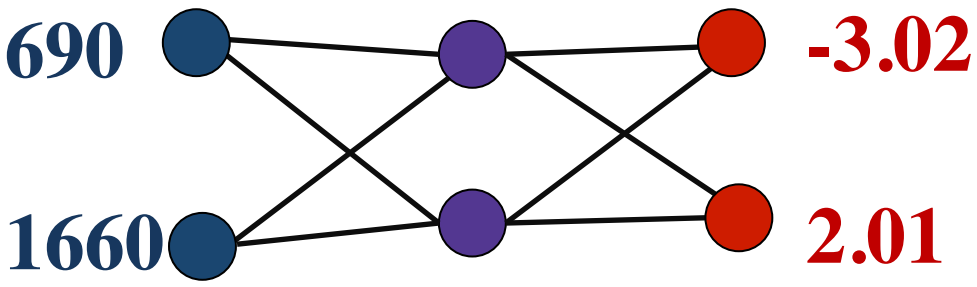
- 各母音の埋め込みベクトル $\phi(v)$ を使って
  点過程のポテンシャルを表現
  - BPP: $\psi(v_i) = |\phi(v_i)|$
  - MPP: $\psi(v_i) = |\phi(v_i)|$
    $\psi(v_i, v_j) = \exp\left(-\frac{\beta}{|\phi(v_i) - \phi(v_j)|^2}\right)$
  - DPP: $p(V) \propto \det\left|\phi(v_1), \cdots, \phi(v_M)\right|$

Deep point process (2)

- 2次元→r次元(特に2次元)の非線形なマッピング
  - 注：特徴は本当は2次元とは限らない
  - 1層のニューラルネット

$$\phi(v) = W_1 \tanh(W_0 f(v) + b_0) + b_1$$

- 層の数は、Devデータで選ぶ {0,1,2,3}

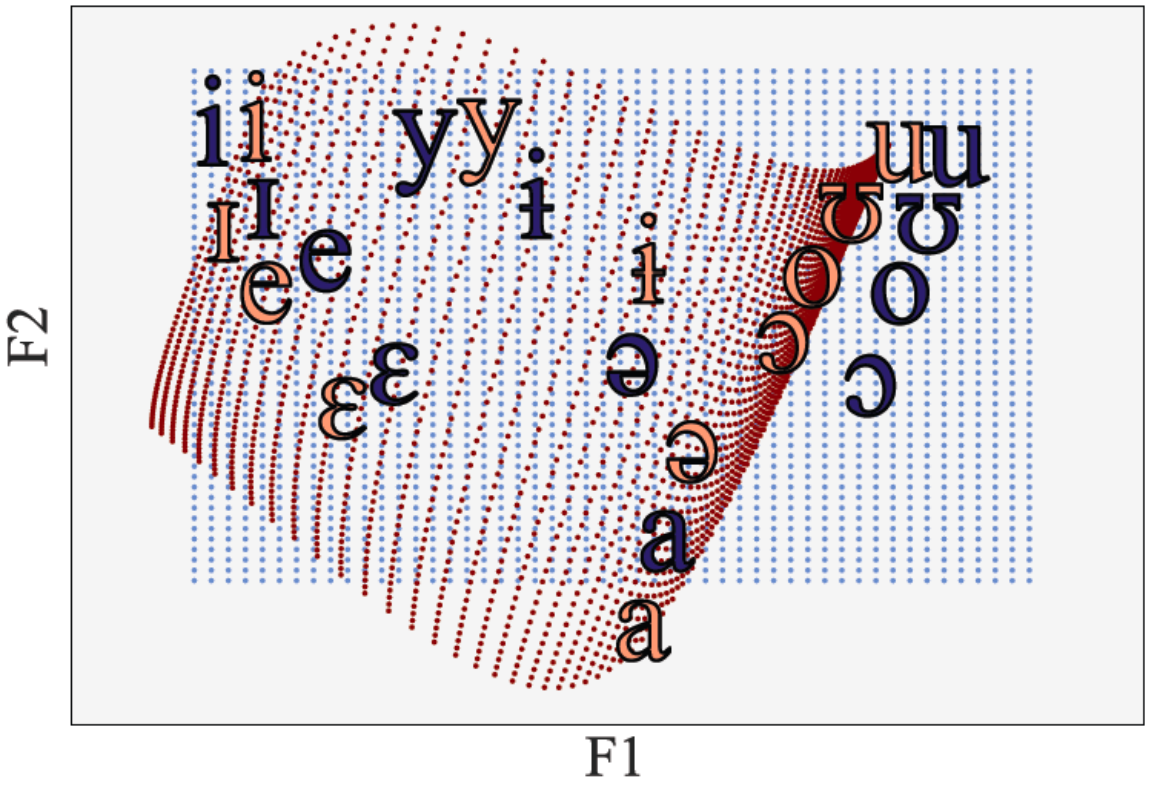


Experiments

- Becker-Kristal (2010) corpus, 223 languages
  - 53個のIPA母音集合で実験
  - 10-fold cross validation (8/10の言語で訓練, 1/10で調節, 1/10の言語でテスト)
- 学習に使わなかった言語の予測精度を評価

	BPP	uBPP	uMPP	uDPP	iBPP	iMPP	iDPP	pBPP	pMPP	pDPP
x-ent	8.24	8.28	8.08	8.00	13.01	11.50	✗	12.83	10.95	10.29
cloze-1	69.55%	69.55%	72.05%	73.18%	64.13%	67.02%	✗	65.13%	68.18%	68.18%
cloze-01	60.00%	60.00%	61.01%	62.27%	61.78%	61.04%	✗	61.02%	63.04%	63.63%
cloze-012	53.18%	53.18%	57.92%	58.18%	39.04%	43.02%	✗	40.56%	45.01%	45.46%

Experiments (2)



- 学習された母音の埋め込み区間 (青:観測フォルマント, 赤: 変換された座標)
  - *r*=3層, プロトタイプ数=20

Discussion

- 母音だけでなく、子音のバリエーションも重要
  - F1,F2だけを特徴とするのは、かなりの簡単化
  - 鼻音, 破裂音などの特徴量を使えるように
- 類型論の問題は、small-data problem
  → 正確な確率モデル化が重要
- actual languages の記述と、natural languages の記述の違い
  - 言語の進化段階を考えると、現在の言語は一定の偏りがある可能性