

# iteR : Un (projet de) package R pour l'analyse de matrices de confusions issues d'expériences de perception phonétique.

Crouzet, Olivier

Laboratoire de Linguistique de Nantes – LLING / UMR6310, Nantes Université / CNRS, France

Rencontres R 2025 – Mons, Belgique – 19-21 Mai 2025



Information / Download

## Introduction : la notion de matrice de confusions

- ▶ Permet de mesurer la qualité d'un système de classification [1, 2, 3].
  - ▶ Entrée / Sortie;
  - ▶ Signal / Réponse;
- ▶ Dans le domaine de la perception de la parole...
  - ▶ Typiquement mises en œuvre dans des tâches N-AFC (choix forcé à N alternatives);
  - ▶ On fait écouter des sons associés à des catégories linguistiques (catégorie d'entrée),
  - ▶ Et on recueille les jugements de catégorisation des auditeurs (réponse catégorielle de sortie);

## Un exemple de matrice de confusions

Cf. Table 1 ci-contre.

## Quantification des résultats

- ▶ **Informations qualitatives** fournies par une matrice de confusions : analyser la nature des erreurs / des confusions;
- ▶ Solutions courantes pour l'**analyse quantitative** des résultats;
  - ▶ Pourcentage de bonnes réponses (diagonale principale, « *accuracy* »), mais ignore la *structure des erreurs* / des confusions;
  - ▶ Diverses mesures issues des travaux en métrologie (sensibilité, spécificité...), notamment accessibles via la bibliothèque caret;
- ▶ Solution quantitative issue de la théorie de l'information : **taux de Transmission d'Information / Information Mutuelle** (SHANNON [1])
  - ▶ Permet d'évaluer la « quantité d'information transmise » (en bits) à travers le canal de diffusion;
  - ▶ Utilisation principes en perception de la parole : MILLER et NICELY [2].
  - ▶ Usages courants dans les travaux sur la perception de la parole (voir par ex. [4, 5])

## Interprétation des valeurs

En perception de la parole, on utilise communément le taux de Transmission d'Information **normalisé** :

Si aucune information transmise (réponses aléatoires)  
 $ITn = 0$ ;

Si information parfaitement transmise (observateur idéal)  
 $ITn = 1$  ou  $ITn = 100\%$  (selon qu'on l'exprime sur 1 ou sur 100);

Entre ces deux extrêmes Degrés variables de transmission d'information / d'Information Mutuelle relative;

## Objectifs

1. Proposer une bibliothèque R qui facilite l'analyse quantitative de matrices de confusions sous l'angle de la théorie de l'information;
2. Notamment adaptée aux travaux en phonétique / perception de la parole;
3. En complément aux méthodes déjà utilisées dans d'autres bibliothèques (caret...);
4. Principal objectif : Mise à disposition de fonctions pour le **calcul du taux d'information transmise / de l'Information Mutuelle**;
5. Fonctions facilitant le regroupement de lignes / colonnes par trait distinctif;

## État d'avancement de la bibliothèque

- ▶ Les principales fonctions de calcul des taux de transfert d'information sont fonctionnelles et leurs résultats ont été vérifiés à partir des données de CHRISTIANSEN, DAU et GREENBERG [5];
- ▶ **À FAIRE** : Mettre à disposition une première version de la bibliothèque accompagnée d'une documentation;
- ▶ **À FAIRE** : Réfléchir à l'apport de méthodes permettant de faciliter la construction des classifications en traits et les implémenter si besoin;

## Références bibliographiques

- [1] C. SHANNON, "A Mathematical Theory of Communication". In : *The Bell System Technical Journal* 27 (1948), p. 399-423, 623-656. — [2] G. A. MILLER et P. E. NICELY, "An analysis of perceptual confusions among some English consonants.". In : *The Journal of the Acoustical Society of America* 27.2 (1955), p. 338-352. — [3] R. M. FANO, *The Transmission of Information : Parts I and II (Technical Reports 65 and 149)*. Rapp. tech. Part I : 34 pages, Part II : 29 pages. Cambridge, Massachusetts : Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT), 1949-1950. — [4] T. U. CHRISTIANSEN et S. GREENBERG, "Perceptual Confusions Among Consonants, Revisited : Cross-Spectral Integration of Phonetic-Feature Information and Consonant Recognition". In : *IEEE Transactions on Audio, Speech, and Language Processing* 20.1 (jan. 2012), p. 147-161. — [5] T. U. CHRISTIANSEN, T. DAU et S. GREENBERG, "Spectro-Temporal Processing of Speech – An information-theoretic framework.". In : *Hearing – From Sensory Processing to Perception*. Springer, 2007. — [6] O. CROUZET, "Perception des consonnes et voyelles nasales en parole vocodée : Analyse de la contribution des niveaux de résolution spectrale et temporelle.". In : *Actes des XXXIèmes Journées d'Etudes sur la Parole – JEP2018*. 4-8 Juin. Aix-en-Provence, France, avr. 2018. — [7] J. WELLS, "SAMPa computer readable phonetic alphabet". In : *Handbook of Standards and Resources for Spoken Language Systems*. Sous la dir. de D. GIBSON, R. MOORE et R. WINSKI. Berlin et New York : Mouton de Gruyter, 1997. Part IV, section B. — [8] K. ISKAROUS et al., "The coarticulation/invariance scale : Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance". In : *The Journal of the Acoustical Society of America* 134.2 (2013), p. 1271-1282. — [9] E. HUME et F. MAILHOT, "The role of entropy and surprisal in phonologization and language change". In : *Origins of Sound Change*. Sous la dir. d'A. C. L. YU. Oxford University Press, jan. 2013, p. 29-48. —

## Traitement de données réelles

Données brutes collectées dans le cadre d'une recherche sur la perception des voyelles avec des simulations d'implants cochléaires [6];

- ▶ Tâche de classification à choix forcé à 13 alternatives;
- ▶ Voyelles isolées;
- ▶ Simulations d'implants cochléaires avec manipulation de différents paramètres acoustiques;
- ▶ Le dataset d'exemple est composé au total de 1170 réponses / observations (dont 3 données manquantes en raison d'une absence de réponse pour les essais concernés), réparties sur 6 auditeurs ayant été soumis à 15 conditions de simulations acoustiques différentes;

## La structure du dataset

Ci-dessous, un extrait aléatoire du dataset restreint à quelques principales colonnes pertinentes (toutes conditions expérimentales confondues, les transcription phonétiques suivent la convention SAMPA (Speech Assessment Methods Phonetic Alphabet, [7]) :

sujet	SAMPA	reponse	success
Axe0	y	y	1
coc5	O	O	1
coc5	2	O~	0
erw3	E	E~	0
erw3	u	u	1
erw3	i	e	0
fra4	O~	a~	0
fra4	a~	9	0
fra4	a	a	1
man2	a~	a~	1

## Obtention de la matrice de confusions complète

- ▶ Pour chaque catégorie d'entrée...
- ▶ On mesure combien de réponses ont été données pour chacune des catégories de sortie;

```
> confusion <- table(dataV$SAMPA, dataV$reponse)
```

Table 1 – Matrice de confusions

	i	e	E	a	y	2	9	u	o	O	E~	a~	O~
i	26	5	0	3	8	1	3	8	18	4	2	1	10
e	22	8	1	5	8	3	8	9	11	6	3	4	2
E	7	13	9	5	4	7	9	3	6	6	8	8	5
a	8	4	4	41	2	5	9	2	1	1	8	2	3
y	22	6	3	5	20	2	1	10	8	4	4	1	4
2	10	0	1	6	4	9	6	10	17	5	6	5	11
9	7	0	6	14	5	10	15	2	9	7	4	4	6
u	6	0	3	7	9	3	12	7	24	6	3	4	6
o	6	2	4	4	3	6	10	4	23	10	7	5	6
O	7	1	3	20	0	14	11	5	8	9	3	4	5
E~	8	2	4	22	3	9	15	3	4	5	6	4	4
a~	7	3	1	19	1	4	14	5	11	6	11	5	3
O~	8	3	0	9	7	10	10	5	15	6	6	7	4

## Génération de matrices de confusions spécifiques

Typiquement, on souhaite travailler sur les confusions en termes de traits linguistiques abstraits (liés à des contrastes phonologiques comme la nasalité, le degré d'aperture...)

- ▶ On a donc besoin de **classifier les données** en fonction de différents critères phonétique / phonologiques;
- ▶ On regroupe les lignes / colonnes en « sur-catégories » **plus abstraites**
  - ▶ ex. 1 : les voyelles nasales vs. les voyelles orales : quel est le taux de transmission d'information du contraste de nasalité?

	+nas	-nas
+nas	50	219
-nas	144	754

- ▶ ex. 2 : les voyelles fermées / moyennes / ouvertes : quel est le taux de transmission d'information du contraste de degré d'aperture?

	close	mid	open
close	160	162	37
mid	169	349	110
open	31	82	67

▶ ...

- ▶ La fonction `factor()` permet assez simplement de générer le codage en termes de traits phonologiques associés;
- ▶ Il faut l'appliquer aussi bien sur la variable d'entrée que sur la variable de sortie;
- ▶ Il ne reste ensuite qu'à utiliser la fonction `table()` pour générer la matrice de confusions associée;

## Formule du taux de Transmission d'Information

- ▶ L'expression du taux d'Information Transmise repose sur le calcul de la co-variance entre chaque catégorie de stimulus et chaque catégorie de réponse.  $T(c)$  est le nombre de bits par trait transmis à travers le canal  $c$ .

$$T(c) = - \sum_{i,j} p_{ij} \log_2 \frac{p_i p_j}{p_{ij}} \quad (1)$$

- $p_i$  Probabilité d'occurrence de chaque classe  $i$  (entrée, probabilité d'observer une ligne = une catégorie en entrée);
- $p_j$  Probabilité d'occurrence de chaque réponse  $j$  (sortie, probabilité d'observer une colonne = une catégorie de réponse);
- $p_{ij}$  Probabilité de co-occurrence d'une catégorie de réponse ( $j$ ) pour une entrée donnée ( $i$ );

- ▶ Cette formule permet d'estimer le **taux de Transmission d'Information « brut », ou « Information Mutuelle »**;
  - ▶ Cette valeur est influencée par un certain nombre de paramètres de l'analyse (nombre de catégories, entropie de l'entrée);
  - ▶ Un calcul supplémentaire permet d'exprimer l'**Information Mutuelle** comme un taux (entre 0 et 1) ou comme un pourcentage (entre 0 et 100) : le **Coefficient d'Incertitude** (qu'on nomme aussi **taux d'Information Transmise normalisé**).
  - ▶ On doit pour celà calculer l'**entropie** de l'entrée (l'incertitude dans les données) :

$$H(i) = - \sum_i p_i \log_2(p_i) \quad (2)$$

- ▶ Finalement, ce qu'on appelle « **taux de Transfert d'Information normalisé, ITn** » (ou Coefficient d'Incertitude) est la division de l'Information Mutuelle par l'Entropie de l'entrée :

$$ITn = \frac{T(c)}{H(i)} = \frac{- \sum_{i,j} p_{ij} \log_2 \frac{p_i \times p_j}{p_{ij}}}{- \sum_i p_i \log_2(p_i)} \quad (3)$$

## Illustration

Pour les données d'exemple restreintes au sous-ensemble le moins dégradé (mais néanmoins représentatif des difficultés que peuvent rencontrer des personnes sourdes portant un implant cochléaire)...

	i	e	E	a	y	2	9	u	o	O	E~	a~	O~
i	4	1	0	1	2	1	0	3	4	1	0	0	1
e	5	2	0	0	3	1	0	2	1	2	0	1	1
E	0	4	0	1	2	2	1	0	3	1	3	1	0
a	0	0	1	11	0	0	4	0	1	0	0	1	0
y	0	1	1	0	12	0	0	0	2	1	0	0	1
2	0	0	1	0	1	1	2	5	5	2	0	0	1
9	0	0	2	3	1	2	3	0	3	3	0	1	0
u	0	0	0	1	1	0	1	4	9	1	0	0	1
o	0	0	0	0	0	1	1	0	9	5	1	1	0
O	0	0	1	1	0	6	2	0	4	4	0	0	0
E~	0	0	0	3	1	4	4	0	1	3	0	1	1
a~	0	0	0	6	0	1	2	0	3	1	4	1	0
O~	0	0	0	0	3	1	4	2	2	3	1	0	2

	+nas	-nas
+nas	10	44
-nas	14	166

	close	mid	open
close	36	34	2
mid	23	90	13
open	0	17	19

- ▶ Pour obtenir le Taux d'Information normalisé (ou « Coefficient d'Incertitude »), on utilise la fonction `iteR::ITnorm()` avec comme argument la matrice de confusions considérée :

```
> res <- iteR::ITnorm(confMatAperture)
> res$ITnorm*100
```
- ▶ La fonction renvoie une liste qui contient différents objets utiles (matrice originale, IT brut, IT normalisé...);
- ▶ Exemple de résultat :
  - ▶ **Pour le contraste de nasalité**, on observe ici un taux de Transmission d'Information normalisé de 1.83%;
  - ▶ **Pour le contraste d'aperture**, on observe ici un taux de Transmission d'Information normalisé de 16%;
  - ▶ Malgré la valeur faible de chacun de ces taux, leur comparaison suggère que le contraste de nasalité est considérablement moins bien transmis / perçu que la différenciation des voyelles sur la base de leur degré d'aperture;
- ▶ On pourrait appliquer ces calculs pour chacune des conditions expérimentales étudiées afin de caractériser leur impact sur la perception des différents contrastes phonologiques;

## Perspectives

- ▶ Élargir l'application à des usages complémentaires en phonétique / phonologie
  1. Sur des données de position des articulateurs consonantiques en fonction de la voyelle contextuelle afin de rendre compte des effets coarticulaires en production (ISKAROUS et al. [8]);
  2. Sur la relation entre probabilité d'occurrence de certaines séquences phonologiques dans une langue et diachronie (changement linguistique au cours du temps, HUME et MAILHOT [9]);