

Théorie de la syllabe et durées vocaliques : Vers une interprétation unifiée du rôle de la structure syllabique et de la nature des segments.

Crouzet, O. & Angoujard, J.-P.

Université de Nantes, Nantes Atlantique Universités,
Laboratoire de Linguistique de Nantes – LLING, EA3827,
UFR Lettres et Langages, Chemin de la Censive du Tertre, BP81227, Nantes, F-44000 France.
{olivier.crouzet|jean-pierre.angoujard}@univ-nantes.fr
<http://www.lettres.univ-nantes.fr/lling/>

ABSTRACT

It is generally agreed that vowel duration may be influenced by both phonetic context and syllable structure. Though it may seem reasonable to call for two different variables in this respect, we show that the *rhythm and substance* [1] approach to syllable structure may offer a common framework for unifying these two sources of variation into a single theoretical account. The results of a speech production experiment involving french speakers are described which confirm that this syllabic approach accounts for some of the predicted deviations in vowel duration when syllabic and contextual effects are involved. Though further studies should be conducted, this framework seems particularly promising for the understanding of the relationship between articulatory, phonological and rhythmic influences on speech production mechanisms.

1. INTRODUCTION

L'étude des facteurs influençant les durées vocaliques fait ressortir un ensemble de variables prédictives complémentaires extrêmement nombreuses. Plusieurs de ces variables sont liées à une représentation prosodique de la chaîne parlée (positionnement de la voyelle dans les groupes intonatifs, structure de la syllabe *portant* la voyelle). L'effet de la nature des segments environnants constitue également un facteur prédictif important des variations observées. Dans la présente communication, nous focalisons notre approche sur une proposition d'unification des modélisations syllabique et segmentale.

1.1. Facteurs influençant les durées vocaliques

L'influence syllabique La structure de la syllabe constitue une source importante de variation de la durée vocalique [10]. Ainsi, la voyelle initiale [a] est plus longue lorsqu'elle est prononcée dans [aʃ] que dans [aʃa]. Dans le premier cas, la fricative est en position terminale (i.e. *coda*, cf. Fig. 1) de la syllabe alors que dans le second cas, elle est en position initiale (i.e. *attaque*) de la syllabe suivante. Maddieson [10] montre que ces variations de la durée vocalique constituent des indices phonétiques (c'est à dire *observables*) de la structure syllabique des séquences. L'influence de la structure syllabique est actuellement prise en considération dans certains systèmes de synthèse vocale [cf. notamment 11].

L'influence segmentale À cette influence syllabique sur les durées vocaliques, s'ajoute une source contextuelle non-prosodique, c'est à dire indépendante de l'organisation *hiérarchique* des segments dans la chaîne parlée :

l'effet du contexte phonétique sur la durée de la voyelle précédente. Dans une séquence VC (Voyelle-Consonne), la voyelle est plus longue si elle est suivie d'une fricative que d'une occlusive (p.ex. [as] vs. [at]) [6, 7], elle est également plus longue si elle est suivie d'une consonne voisée que d'une consonne non-voisée (p.ex. [ad] vs. [at]) [3]. Ces phénomènes pourraient cependant être modulés par, voire confondus avec, d'autres variables comme la vitesse d'élocution. Dans les mécanismes de perception de la parole, ces variations contextuelles peuvent être utilisées comme des indices acoustiques du voisement des consonnes post-vocaliques. Lindblom [9] rend compte de ces variations à partir d'un modèle de la coordination des lèvres et de la mâchoire.

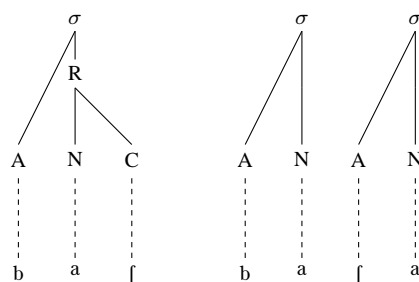


FIG. 1: Représentation syllabique de [baʃ] et [baʃa] selon un cadre théorique classique communément utilisé dans les travaux actuels sur la syllabe. σ représente le *sommet* de la syllabe. R désigne la *rime*, elle-même constituée du *noyau* (N) et de la *coda* (C). L'*attaque* (A) est directement rattachée au sommet de la syllabe.

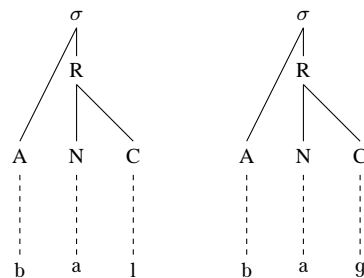


FIG. 2: Représentations syllabiques de [bal] et [bag] dans le cadre théorique classique. La représentation des séquences [bal] et [bag] est strictement identique dans ce cadre-là.

2. LA SYLLABE : RYTHME ET SUBSTANCE

Dans le cadre de la théorie syllabique développée en [1], la syllabe est considérée comme l'instanciation d'une chaîne de segments et d'une grille rythmique (cf. Fig. 3).

Tout segment est caractérisé par un ensemble de propriétés qui déterminent sa nature, c'est à dire sa *substance* [4, 8] et se voit attribuer une position rythmique déterminée à partir de ses caractéristiques substantielles. Toute voyelle est un pic rythmique, toute consonne est un creux rythmique. Ces creux rythmiques peuvent avoir des statuts différents. En autorisant le rattachement d'un segment donné à une position de creux post-syllabique, on lui attribue le statut de *coda* (ici appelée « Position 3 »). Seule une classe particulièrement limitée de segments peut se rattacher à cette position 3 (notamment le /l/). La plupart des segments consonantiques se rattachent nécessairement à une « Position 1 » (c'est à dire à une attaque de syllabe). Le statut des fricatives (et par conséquent du /ʁ/ français) est à ce titre encore sujet à débat dans le cadre de cette approche.

L'interaction entre nature des segments et grille rythmique détermine les caractéristiques de la courbe supérieure, laquelle est en général interprétée comme une représentation de la sonorité [5] mais pourrait également être vue comme une représentation symbolique des alternances d'ouverture et de fermeture du conduit vocal.

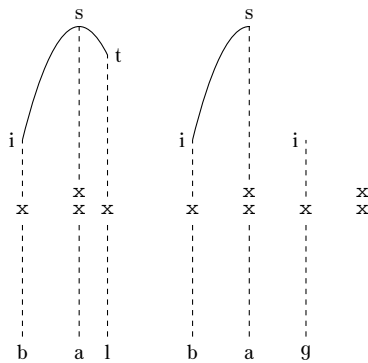


FIG. 3: Représentation syllabique de [bal] et [bag] selon [1]. La ligne inférieure correspond à la séquence des segments ; sur la ligne médiane, on trouve la grille rythmique qui représente l'alternance des position rythmiques (faibles / fortes) de la séquence ; la ligne supérieure (la courbe) est classiquement décrite comme une représentation de l'échelle de sonorité des segments. Ces trois niveaux de représentation fournissent une description particulièrement riche de ce qu'est une syllabe dans ce cadre théorique. Contrairement aux conceptions classiques, [bal] et [bag] ont une représentation syllabique toute à fait différente : [g] ne peut pas être rattaché en position finale de syllabe et se trouve nécessairement à l'attaque d'une syllabe à noyau vide.

Dans ce cadre théorique, les séquences [bal] et [bag] se voient attribuer une représentation profondément différente. Cette différence est du même ordre que celle qui caractérise l'opposition entre les formes [aʃ] et [aʃa] : dans le premier cas, la consonne est en position finale de syllabe, dans le second cas elle se trouve à l'attaque de la syllabe suivante. Ce parallélisme permet de faire des prédictions concernant les variations de durée vocalique en fonction de la structure syllabique *et* de la nature des segments.

Si les représentations syllabiques présentées ici permettent de prédire la réalisation de voyelles plus courtes dans [bal] que dans [bag], c'est que l'on peut voir dans ces représentations syllabiques une conception proche des propositions d'Öhman [12] concernant la coordination des gestes vocalique et consonantique lors de la production de la parole [cf. également 2]. En effet, selon [12] le geste vocalique est un geste *global* ; tout geste consonan-

tique n'est quant à lui qu'un geste *local* et transitoire entre deux gestes vocaliques. Si l'on intègre ces propositions dans la théorie syllabique décrite ici, on peut considérer qu'un geste vocalique s'étale sur toute la largeur de la syllabe alors que les gestes consonantiques auront une portée beaucoup plus réduite. Nous proposons que ces gestes vocaliques sont limités à l'empan de la syllabe (en tant que représentation symbolique de la chaîne parlée). Passer d'une syllabe à une autre consiste à passer d'une voyelle à une autre en traversant des états consonantiques, ces états consonantiques étant rattachés à la position initiale ou finale d'une syllabe en fonction de leur nature. Ces propositions sont représentées dans les Figures 4 à 6.

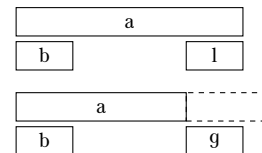


FIG. 4: Représentation temporelle de [bal] et [bag]. Le modèle syllabique développé en [1] permet de proposer une représentation temporelle proche des propositions de [12]. Cette représentation conduit à prédire des différences de durée vocalique en fonction de la nature des segments. Le modèle prédit des variations de la durée des voyelles en fonction de la structure syllabique sur la base des mêmes principes (cf. Fig. 5 et 6).

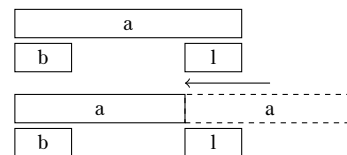


FIG. 5: Représentation temporelle de [bal] et [bala]. La diminution de la durée de la voyelle dans [bala] s'explique par la réduction de l'empan attribué à la voyelle dans cette forme.

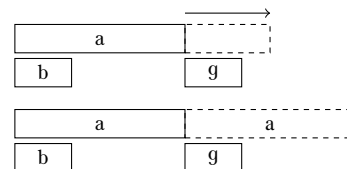


FIG. 6: Représentation temporelle de [bag] et [baga]. L'augmentation de la durée de la voyelle dans [baga] peut s'expliquer ici par l'allongement de l'empan global attribué au geste vocalique (incluant donc aussi la portion correspondant au noyau vide).

3. EXPÉRIENCE

Nous faisons l'hypothèse que la conception rythmique de la syllabe est un modèle particulièrement approprié pour rendre compte des phénomènes phonologiques associés à la syllabe mais aussi des phénomènes articulatoires et acoustiques qui pourraient interagir avec cette représentation.

Afin de valider cette hypothèse, nous avons étudié les variations de la durée des voyelles en fonction de leur contexte consonantique *et* de la structure syllabique prédite par les deux approches théoriques présentées. En effet, si le modèle syllabique classique conduit à considérer les effets syllabiques et contextuels comme deux effets indépendants, l'interaction forte entre représentation syllabique et segmentale mais aussi entre représentation phonologique et contraintes motrices est centrale dans le

modèle rythmique. À partir de cette opposition constitutive des deux modèles, il est possible de faire l’hypothèse suivante. Si l’on compare par exemple les sons [l] et [g], le modèle rythmique prédit une représentation syllabique fondamentalement différente sur la base de leur nature segmentale, ce que ne fait pas le modèle syllabique classique. Dans ce cadre-là, le modèle rythmique prédit des différences de durée vocalique entre les voyelles produites dans [bal] et [bag], ce qui peut évidemment s’expliquer, en dehors du modèle syllabique, par des contraintes articulatoires telles que celles décrites par Lindblom [9].

Mais le modèle rythmique permet également de faire une prédiction tout à fait centrale : si la consonne finale du mot est effectivement rattachée à la syllabe suivante dans des séquences comme [labaletõbe] ou [labagetõbe], les représentations des consonnes qui nous intéressent sont quant à elles particulièrement différentes lorsqu’elles sont suivies d’une autre consonne ne donnant pas lieu à resyllabation. Ainsi, les mots [bal] et [bag] dans [labaldõmõfɛɛɛ] et [labagdõmõfɛɛɛ] ne peuvent-ils être analysés comme structurellement identiques. Cette différence a un impact majeur sur les prédictions que l’on peut faire concernant les durées vocaliques. En effet, s’il est probable que l’on observera des différences de durée entre les séquences de type CV#CV (où # représente une *frontière* syllabique dans le cadre du modèle syllabique classique) et celles de type CVC#C, le modèle rythmique –contrairement au modèle classique– prédit que cet effet devrait être plus fort pour la latérale alvéolaire [l] –qui changera effectivement de position syllabique– que pour les occlusives voisées – qui elles ne peuvent jamais être rattachées à une position de coda, Cf. Fig. 7 et 8 ainsi que Fig. 3.

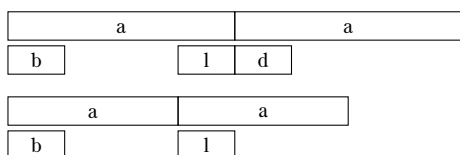


FIG. 7: Représentation temporelle de [balda] et [bala]. La resyllabation du [l] conduit à prédire une réduction particulièrement importante de la durée de la voyelle dans [bala].

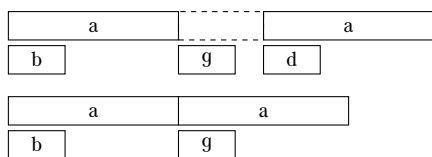


FIG. 8: Représentation temporelle de [bagda] et [baga]. L’absence de chevauchement entre les gestes vocalique et consonantique final dans [bag] conduit à prédire un fonctionnement totalement différent du passage de [bagda] à [baga]. En effet, dans les deux cas, le geste vocalique ne se prolonge pas dans la consonne suivante en raison de la présence d’un noyau vide (cf. Fig. 3). Les variations de durée devraient donc être moins importantes.

3.1. Méthode

Participants Nous présentons ici les résultats de deux locuteurs ayant participé à cette expérience. Nous procédons actuellement à l’enregistrement d’un nombre plus conséquent de locuteurs mais présentons néanmoins les résultats des analyses statistiques inférentielles conduites sur ces données afin de mettre en évidence la stabilité des résultats observés. Les locuteurs sont tous de langue maternelle française et ne présentent aucun trouble avéré de la production de la parole.

Enregistrement Les enregistrements ont été réalisés sur la piste gauche d’un enregistreur DAT Tascam DA-P1 avec un microphone Sennheiser e835. Les locuteurs étaient assis face au microphone dans une pièce calme et lisaient la liste des phrases à un rythme d’une phrase toutes les 2 secondes. Les enregistrements ont ensuite été transférés sur ordinateur par l’intermédiaire d’une carte son Sound-Blaster Audigy LS et digitalisés en monophonique sur 16 bits à 16kHz.

Matériel Le noyau syllabique des mots étudiés correspondait à l’une des 4 voyelles {i,a,y,u} (3 mots différents par voyelle * 6 classes de consonnes post-vocaliques : [ʁ] vs. [l] vs. {v,z,ʒ} vs. {f,s,ʃ} vs. {b,d,g} vs. {p,t,k}). Il s’est avéré impossible de maintenir la consonne initiale constante pour chaque *rime* VC ; les mots ont donc été choisis afin de s’assurer d’un contrebalancement des caractéristiques de la consonne initiale, ceci afin de limiter les effets éventuels de cette consonne sur l’analyse des durées vocaliques : pour chaque ensemble de mots se terminant par une classe de consonnes donnée, on s’est assuré d’une répartition équitable d’occlusives, de fricatives et de nasales, d’une part ; de voisées et de non-voisées d’autre part. Au total, chaque locuteur a produit les 72 mots dans 2 contextes différents, soit 144 réalisations par locuteur.

Mesures acoustiques La mesure des durées vocaliques a été réalisée à l’aide du logiciel Wavesurfer [14] à partir de la courbe de modulation d’amplitude et du spectrogramme à bande large en réglant le contraste pour maximiser l’émergence des formants vocaliques. Afin de réduire la variabilité des mesures effectuées, nous avons choisi *a priori* un ensemble de critères de localisation du début et de la fin d’une voyelle. Ces critères sont les suivants :

- Lorsque la consonne précédente (resp. suivante) est non-voisée, le début (resp. la fin) de la voyelle est localisé au moment où apparaissent (resp. disparaissent) *simultanément* les formants et l’énergie correspondant au voisement.
- Lorsque la consonne précédente (resp. suivante) est voisée, le début (resp. la fin) de la voyelle est localisé au moment où apparaît la frontière entre bruit de frottement (lié à la friction ou au relâchement) et tracés formantiques (resp. au moment où commence la phase d’occlusion ou de friction).
- En cas d’incertitude, nous avons fait reposer notre décision sur l’écoute du signal et c’est toujours la solution qui tend à réduire la durée de la voyelle qui a été privilégiée.

Si ces critères peuvent de toute évidence être sujet à débat (comme probablement tout critère dans l’analyse des données scientifiques), leur respect nous permet d’atteindre un maximum de régularité dans les décisions, ce qui nous paraît être le plus important pour l’étude de cette hypothèse d’interaction en fonction du contexte et de la structure.

3.2. Résultats

Le Tableau 1 présente les durées moyennes de la voyelle en fonction de la nature de la consonne post-vocalique et du contexte dans lequel elle est produite.

TAB. 1: Durées vocaliques mesurées (en ms.) en fonction de la nature de la consonne finale et de son contexte de production (contexte vocalique donnant lieu à resyllabation vs. contexte consonantique ne donnant pas lieu à resyllabation).

Consonne post-vocalique	Structure CV#CV	Structure CVC#C
ʁ	93.7	94.5
l	97.2	89.7
v,z,ʒ	108.2	101.0
f,s,ʃ	78.4	84.8
b,d,g	75.2	73.3
p,t,k	73.6	71.5

Nous avons conduit une analyse de variance de ces données en évaluant les effets de la nature de la voyelle, de la consonne post-vocalique et de la structure syllabique provoquée par le contexte post-consonantique sur la durée vocalique. On n’observe aucune interaction du timbre de

la voyelle avec les autres facteurs ($F < 1$). L'interaction de 2nd ordre est marginale ($F_{[30,60]} = 1.56, p = 0.07$). Cette variable a donc été retirée des analyses ultérieures. Les effets globaux du type de consonne post-vocalique et de la structure syllabique déclenchée par le contexte de production du mot sont significatifs (respectivement $F_{[5,8]} = 64.20, p < 0.01$ pour l'effet du contexte et $F_{[2,8]} = 14.24, p < 0.01$ pour la structure syllabique). L'interaction entre ces deux variables est également significative ($F_{[10,20]} = 22.29, p < 0.01$).

Nous avons vu rapidement, dans l'introduction (Sec. 2), que les fricatives et le [ʁ] soulèvent de nombreuses questions quant à leur représentation dans le modèle rythmique. Nous nous limiterons donc ici à une analyse comparative de la latérale alvéolaire ([l]) et des occlusives voisées ({b,d,g}) afin d'évaluer la validité des hypothèses formulées (nous laissons de côté les occlusives non-voisées pour ne pas introduire dans la comparaison de modification liée au voisement). À cet effet, nous avons conduit une analyse des contrastes pertinents pour répondre à la question suivante : la différence de durée vocalique observée en fonction de la structure syllabique (CV#CV vs. CVC#C) est-elle plus importante pour la consonne ([l]) que pour les occlusives voisées (les données spécifiquement associées à cette question sont illustrées Fig. 9) ?

En structure CV#CV, les voyelles sont plus longues lorsqu'elles sont suivies d'une latérale alvéolaire que d'une occlusive voisée ($p < .05$). En structure CVC#C, on observe une différence de durée vocalique plus importante, laquelle est également significative ($p < .001$).

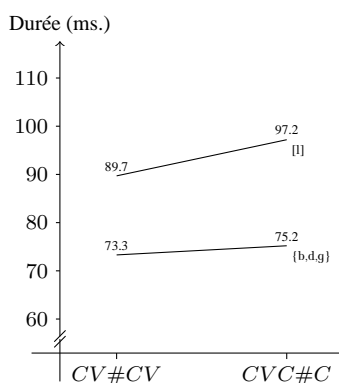


FIG. 9: Comparaison des durées vocaliques mesurées (en ms.) en fonction de la consonne post-vocalique ([l] vs. {b,d,g}) et de son contexte de production (contexte vocalique donnant lieu à resyllabation vs. contexte consonantique ne donnant pas lieu à resyllabation).

4. DISCUSSION

S'il n'est pas possible d'affirmer que nous sommes ici en présence d'une *interaction statistique* entre contexte phonétique de la voyelle et structure syllabique, le fait d'aboutir à un seuil de probabilité nettement plus bas pour les structures CVC#C que pour les structures CV#CV nous semble particulièrement prometteur. En effet, à la lecture de ces analyses, il semble que l'allongement de la durée de la voyelle en fonction de la structure syllabique puisse effectivement être plus important devant [l] que devant une occlusive voisée puisque le test statistique fait *plus facilement* émerger cet effet.

Il conviendra de poursuivre ce travail en comparant diffé-

rents modes de calcul des durées vocaliques ainsi qu'en ayant systématiquement recours à une transformation des données de durée segmentale en distribution log-normale [13] afin d'accroître la validité statistique des analyses mais aussi de recourir à des statistiques Bayésiennes afin d'évaluer à l'aide d'outils statistiques plus appropriés les modifications d'*ampleur* des effets observés.

Quoi qu'il en soit, les résultats présentés ici constituent une première tentative d'unification de deux sources de variation des durées vocaliques considérées classiquement comme deux phénomènes distincts. Des travaux complémentaires doivent évidemment être conduits afin d'approfondir notre compréhension de ces phénomènes. Cette approche théorique nous semble cependant particulièrement intéressante, et les résultats préliminaires présentés dans cette étude contribuent à renforcer cette position, car elle nous semble constituer un outil conceptuel tout à fait approprié à une réflexion sur les relations entre contraintes articulatoires, représentations phonologiques et mécanismes cognitifs de type rythmique.

RÉFÉRENCES

- [1] J.-P. Angoujard. *Théorie de la syllabe*. Paris : CNRS Éditions, 1997.
- [2] C. P. Browman and L. Goldstein. Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston and M. E. Beckman, editors, *Papers in Laboratory Phonology I : Between the grammar and physics of speech*, pages 341–376. Cambridge University Press, Cambridge, UK, 1990.
- [3] M. Chen. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22 :129–159, 1970.
- [4] G. N. Clements. The geometry of phonological features. *Phonology Yearbook*, 2 :225–252, 1985.
- [5] G. N. Clements. The role of the sonority cycle in core syllabification. In J. Kingston and M. E. Beckman, editors, *Papers in laboratory phonology I : Between the grammar and physics of speech*, pages 283–333. Cambridge University Press, Cambridge, UK, 1990.
- [6] A. House. On vowel duration in English. *The Journal of the Acoustical Society of America*, 33 :1174–1178, 1961.
- [7] A. House and G. Fairbanks. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25 :105–113, 1953.
- [8] J. Kaye, J. Lowenstamm, and J.-R. Vergnaud. The internal structure of phonological elements : A theory of charm and government. *Phonology*, 7(2) :193–231, 1990.
- [9] B. Lindblom. Vowel duration and a model of lip mandible coordination. In *Speech Transmission Laboratory – Quarterly Progress and Status Report*, volume 8, pages 1–29. Royal Institute of Technology, Stockholm, 1967. URL http://www.speech.kth.se/qpsr/pdf/1967/1967_8_4_001-029.pdf.
- [10] I. Maddieson. Phonetic cues to syllabification. In V. Fromkin, editor, *Phonetic Linguistics : Essays in Honor of Peter Ladefoged*, pages 203–221. Academic Press, CA, Orlando, 1985.
- [11] R. Ogden, J. Local, and P. Carter. Temporal interpretation in prosynth, a prosodic speech synthesis system. In *Proceedings of the XIVth International Congress of Phonetic Sciences, ICPH'99*, 1999.
- [12] S. E. G. Öhman. Coarticulation in VCV utterances : Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39 :151–168, 1966.
- [13] K. M. Rosen. Analysis of speech segment duration with the lognormal distribution : A basis for unification and comparison. *Journal of Phonetics*, 33 :411–426, 2005.
- [14] K. Sjölander and J. Beskow. Wavesurfer : An open-source speech tool. In *International Conference on Spoken Language Processing*, Beijing, China, 2000. URL <http://www.speech.kth.se/wavesurfer/>.