

Le poids perceptif de la durée de la voyelle face à sa qualité acoustique dans la discrimination des voyelles longues et brèves

Omran Allatif

GIPSA-Lab - Institut de la Communication Parlée -UMR 5216
CNRS/INPG/UJF/Stendhal
Université Stendhal- BP 25 - 38040 Grenoble cedex 9, FRANCE
Tél. : ++33 (0)4 76 82 41 20 - Fax : ++33 (0)4 76 82 43 35
Omran.Allatif@icp.inpg.fr

ABSTRACT

This study investigated the vowel duration's contribution in the vowel identification when the vowel is a part of a Short vs. Long two terms contrast. Listeners were identifying modified and non-modified natural speech sounds; Stimuli were the 3 cardinal long vowels, and their 3 short partners, of the Arabic dialect of Mayadin (Syria). In this experiment long vowels were shortened to reach the natural duration of shorts, short ones underwent lengthening to reach the natural duration of long ones. Results showed that durational opposition does not have the same importance for each of the three pairs. Effectively, when temporal contrast was removed, the /a/ and /a:/ identities were confused at 80%, /u/ and /u:/ were confused at 47%, while /i/ and /i:/ they were confused at 18% only. The findings support the hypothesis that vocalic categories don't rely upon the same information nature to make opposition with their quantitative contrastive partners.

1. INTRODUCTION

De nombreuses études se sont intéressées à la mesure de la durée et/ou de l'espace vocalique en Arabe ([Ani70]; [Hai94]; [New02]; [Bar03]; [All04]). Les études qui ont porté sur la durée ont montré une différence importante entre la voyelle brève et sa partenaire longue. [Bar03] montre que le rapport longue/brève est de 1,9 pour l'Arabe du Maghreb, alors qu'il est de 2,3 pour les régions du Moyen-Orient. [All04] ont trouvé que ce rapport était de 1,7 pour la région de l'Euphrate en Syrie. Les études portant sur la qualité vocalique conviennent que les voyelles brèves se situent dans une zone centrale de l'espace vocalique par rapport aux voyelles longues, les travaux de [New02] sur l'Arabe du Caire vont dans ce sens aussi, [Abo94] avait établi l'espace vocalique de 8 dialectes du monde arabe; elle confirme la tendance à la centralisation pour tous les dialectes qui ont fait l'objet de son étude.

La question de la prédominance des traits phonétiques "quantitatif vs. qualitatif", parmi les indices acoustiques, reste mal évaluée dans les études précédentes, en effet, nous ne savons pas si toutes les catégories fondent leurs oppositions quantitatives sur l'indice temporel au même point pour démarquer les longues des brèves. Par

ailleurs, les distances perceptives entre les voyelles dans les espaces considérés ne sont pas étudiées. Ainsi, aux voyelles brèves, les phonéticiens assignaient systématiquement la qualité de la longue, en considérant que la différence se situe uniquement au niveau de la durée.

Cette question de la prédominance "Qualité vs. Quantité" était discutée pour d'autres langues, mais pour l'Arabe. En effet, dans une étude perceptive avec des voyelles synthétisées, [Ain72] a montré qu'en anglais britannique la durée affecte l'identité de la voyelle. Il affirme que certaines voyelles, proches en termes de F1-F2, notamment celles du centre de l'espace vocalique, comme /ɜ/-/ʌ/ et /o/-/ɔ,ɑ/ peuvent être confondues entre elles si le jugement de l'identité de la voyelle n'est basé que sur les informations acoustiques spectrales, c'est-à-dire que la quantité est dominante. [Mer78] avait montré qu'en Anglais américain les deux voyelles / / et /æ/, insérées dans des syllabes synthétisées /bVd/ et /bVt/, en raison de leurs durées très différentes et de leurs qualités acoustiques proches, elles sont identifiées en fonction de leurs durées, la longue est perçue /æ/ alors que la brève est perçue /ɛ/.

Le problème était aussi bien posé pour le Français où le contraste temporel n'est pas phonologique; Effectivement, [Got88] ont mené une étude qui a examiné l'impact de l'information temporelle vs. spectrale dans la discrimination des deux voyelles de la paire /o/-/ɔ/, soit la paire minimale /kot/ (côte) vs. /kɔt/ (cotte). Ces deux voyelles présentent des différences et au niveau quantitatif et au niveau qualitatif. [Got84] ont montré que /o/ est 25% plus longue que /ɔ/ dans un contexte /tVt/. Leurs résultats pour des natifs du Français montrent que la durée n'est pas utilisée comme un indice pour identifier la voyelle, les sujets sont insensibles à l'information temporelle qui sera insuffisante pour démarquer les 2 voyelles, ici l'information spectrale est dominante. Lorsque cette même expérience était conduite auprès des natifs de l'Anglais des États-Unis, la catégorisation était influencée par la durée de la voyelle du Français. En effet, lorsque l'information spectrale devient ambiguë, lors du passage de /o/ à /ɔ/ les valeurs formantiques pouvant se situer entre les deux qualités, alors les locuteurs natifs de l'Anglais américain, ceux qui connaissent le Français et ceux qui ne le connaissent pas identifient la longue des deux comme /o/ et la brève en

tant que /ɔ/, dans ce cas, l'information temporelle est suffisante pour démarquer les deux identités.

Contrairement aux études précédentes, ici, nous exposons un procédé expérimental, "la neutralisation de trait", dont les limites et le bien-fondé restent à tester en l'appliquant à d'autres langues présentant, comme l'Arabe, des contrastes temporels phonologiques, et dont le but est de sonder, uniquement, l'étendue du rôle de la durée de la voyelle comme un indice acoustique face à un autre indice qui pourrait avoir un rôle primordial dans cette opposition, à savoir, les valeurs spectrales des formants. Le rôle de la durée de la syllabe, dont la durée des unités peut être aussi utilisée en tant qu'indice pour la durée de la voyelle, sera discuté.

2. METHODE

2.1. Stimuli

Les mesures de la durée des voyelles du système de l'Arabe de Mayadin (Syrie), ainsi que leurs positions dans l'espace vocalique en termes de F1 et de F2, ont été déjà établies [All04]. Nos mesures donnaient les valeurs suivantes (moyennes en ms) : /i:/=196; /a:/=237 et /u:/ = 209, les correspondantes brèves : /i/=120, /a/=143 et /u/=125.

Notre méthode expérimentale a consisté en la neutralisation de l'effet de la durée de la voyelle, parce que nous supposons que si nous éliminons la différence temporelle entre voyelles et que l'opposition se maintient à 100%, c'est que la durée n'est pas considérée comme un indice, alors que si l'opposition disparaît à 100%, c'est que la durée est le seul indice qui se tient derrière cette opposition. La neutralisation s'est faite pour les 3 longues et pour les 3 correspondantes brèves. Les 6 voyelles, qui ont fait l'objet de notre étude, étaient portées par des phrases de la parole naturelle, prononcées par un locuteur natif du dialecte avec un débit normal de la parole combiné avec un schéma mélodique déclaratif. La F0 était de 110 Hz. Chaque phrase était formée de deux mots monosyllabiques CVC CVC, les deux sont lexicalisés "gal zVr" (il a dit zVr) où V pouvait être l'une des 6 voyelles citées ci-dessus. À chaque voyelle différente correspondait un sens différent du mot "zVr" [All04].

Signaux de départ : Il y avait 10 signaux de base pour chaque voyelle, lesquels étaient 10 émissions différentes pour la phrase CVC CVC par un même locuteur. Ces signaux ont été enregistrés dans une chambre sourde sur un PMD 670 et échantillonnés à 44100 Hz.

Signaux cible : chaque signal de base s'est vu modifier à son tour, une seule fois, la voyelle de la deuxième syllabe, uniquement la voyelle, le signal de base modifié était alors présenté pour l'identification, lequel avait une seule occurrence par locuteur. Après modification, nous avons une liste de signaux à identifier, la liste contenait 20 signaux, 10 signaux de base (départ) et 10 autres

modifiés (cibles), pour chacune des 6 voyelles, pour un total de $20 \times 6 = 120$ stimuli à identifier par locuteur. Finalement, nous avons $20(\text{signaux}) \times 6(\text{voyelles}) \times 10(\text{sujets}) = 1200$ identifications. Introduire autant de signaux modifiés que des signaux non modifiés sur la liste des phrases à identifier avait pour but de savoir si le sujet avait un problème avec la reconnaissance du contraste quantitatif dans la parole naturelle au moment de l'identification des mots.

2.2. Sujets

Les sujets étaient de nombre de 10, tous volontaires, 4 hommes et 6 femmes âgés de 18 à 42 ans (la moyenne = 30 ans), tous natifs du dialecte et tous connaissent personnellement la personne qui a prononcé les signaux de base et ont bien l'habitude de l'entendre parler, la plupart exercent des métiers intellectuels. Ils sont naïfs en ce qui concerne la phonétique. Aucun ne manifestait des troubles psychologiques ou des problèmes de l'audition.

2.3. Procédure

Nous avons essayé, lors de la manipulation des signaux, de laisser intacts les indices qui serviraient à l'identification de la qualité de la voyelle (les transitions formantiques, les valeurs des formants, la F0 et l'intensité). En effet, la modification des signaux n'a porté que sur la dimension temporelle dans la partie stable de la voyelle. Notre matière première était formée de 10 signaux de base pour chacune des six voyelles, tous différents, lesquels étaient dupliqués et ensuite intégrés sur la liste des phrases à identifier comme des signaux contrôle. C'étaient les signaux qui étaient des copies des signaux de base qui ont subi la modification. Nous étions face à deux cas de figure :

1) La voyelle est brève: nous commençons par déterminer la durée de la voyelle en question, parce que l'allongement, comme le raccourcissement, se faisait proportionnellement à la durée initiale. Nos mesures précédentes ont montré que pour les trois catégories vocaliques [i a u], la durée de la brève constituait $\approx 0,60$ % de la durée de la longue correspondante. Ainsi, pour savoir de combien il fallait allonger la brève, nous divisons sa durée initiale par 0,6. Le résultat était la valeur de la durée que nous devons atteindre. Garder la proportionnalité permet de produire la même variabilité de longueurs chez les signaux modifiés que la variabilité observée auprès des signaux naturels. L'allongement se faisait par copier un cycle délimité par deux passages par zéro dans le début la partie stable de la voyelle ($T \approx 9$ ms), et par le reproduire juste après le cycle original. Cette opération était répétée pour deux cycles sur trois jusqu'à atteindre la valeur de: "durée initiale/0,6", nous nous arrangeons, en supprimant ou en rajoutant un cycle, par rapport à ce qui était prévu, pour que la durée atteinte coïncide au mieux avec la durée souhaitons atteindre.

2) La voyelle est longue: la procédure était la même que celle suivie pour la voyelle brève, sauf qu'ici, chaque voyelle s'est vue perdre 40 % de sa durée initiale, cela s'effectuait par couper un cycle sur trois, puis un cycle sur deux, puis un cycle sur trois et ainsi de suite jusqu'à atteindre 60 % de la durée initiale de la voyelle objet de la modification. Ces

signaux étaient manipulés avec le logiciel Praat, version 4.0.49, et lus, lors de l'identification, par la même version du même logiciel.

2.4. Identification

Les signaux, modifiés et non modifiés, étaient mélangés ensemble et présentés dans un ordre aléatoire. Le sujet pouvait écouter la phrase autant de fois qu'il voulait. Face à chaque phrase de la liste, le locuteur avait deux possibilités: soit la phrase était claire auditivement et là il identifiait et écrivait le deuxième mot de la phrase "zir, zi:r, zar, za:r, zur ou zu:r". Soit la phrase n'était pas claire, parce que le deuxième mot n'était pas distinct; Ici, il cochait une case prévue pour cette fin; Cette deuxième possibilité était introduite afin de détecter les signaux qui auraient pu être mal synthétisés. Le locuteur prenait le temps dont il avait besoin pour noter le mot qu'il percevait sans souci de vitesse ni de performance.

3. RESULTATS

3.1. Signaux non modifiés

(table 1): Les résultats du test perceptif montrent que certaines parmi les voyelles qui n'ont subi aucune manipulation ont été soit mal identifiées, soit considérées "pas claires". Les longues /i:/, /a:/ et /u:/ sont toutes robustes, leur taux d'identification correcte est de 98%. Ordre de robustesse : 1) /i:/. 2) /a:/ 3) /u:/.

La surprise vient du côté des brèves, surtout le /u/ qui est pris 19 fois sur 100 pour le long correspondant. Le /a/ est considéré comme une voyelle qui prêtait à confusion et difficile à reconnaître. C'est le /i/ bref qui se maintient le mieux parmi les brèves, il est rarement confondu avec la longue et rarement indistinguable. L'ordre de robustesse des voyelles brèves dans la parole naturelle est le même que celui des longues, à savoir 1) /i/. 2) /a/ et 3) /u/.

Les fausses identifications étaient réparties sur la quasi totalité des sujets, ce qui a rendu injustifiable l'élimination de l'un ou de l'autre.

Table 1: Résultats du test perceptif pour les signaux non modifiés.

Voyelles non modifiées	Identification correcte	Identification Fausse	Signal pas clair
/a/	87	3	10
/a:/	98	0	2
/i/	94	4	2
/i:/	100	0	0
/u/	77	19	4
/u:/	95	2	3

3.2. Signaux modifiés

(table 2) : Nous considérons que la modification était réussie lorsque le sujet identifiait sans ambiguïté le signal de base transformé en tant que signal cible. Par exemple, un /i/ transformé en /i:/ et identifié en tant que /i:/ était une réussite, alors qu'un /i/ transformé en /i:/ et identifié en tant que /i/ était un échec.

Nous constatons que les longues, toutes catégories confondues, sont plus intelligibles, comme elles maintiennent mieux leurs identités sous la modification. Nos résultats montrent que c'est la catégorie /a/ qui est la plus sensible au changement de la durée. Effectivement, 81% des /a/ allongés étaient identifiés en tant que /a:/, et 79 % des /a:/ raccourcis étaient identifiés en tant que /a/, cela nous laisse penser que l'opposition au sein de cette catégorie compte essentiellement sur la durée. Avec un taux de réussite inférieure à 20%, la catégorie /i/ est de loin la catégorie la moins sensible à l'effet de la durée. Quant à la catégorie /u/, son taux de réussite s'établissant à 47%, se situe quasiment à égale distance des deux autres catégories.

Si nous classons les voyelles en termes de la préservation de l'identité sans l'effet de la durée, nous trouverons l'ordre suivant /i:/, /u:/, /i/, /u/, /a:/ puis /a/.

Nous pensons que la mauvaise identification des signaux non modifiés, et la hausse du taux de l'ambiguïté pour les signaux modifiés, provient du fait que les locuteurs s'attendaient à ce qu'il y ait des stimuli ambigus, puisque "identifier les signaux équivoque" faisait partie de la tâche. Il est possible que si la consigne en était différente, où les sujets n'avaient que deux choix forcés "brève ou longue", au lieu de 3 "brève, longue ou équivoque", les résultats quant aux voyelles non modifiés seraient différents de ce que nous avons trouvé.

Table 2: Résultats du test perceptif pour les signaux modifiés.

Voyelles modifiées	Réussite de la Modification	Échec de la modification	Signal pas clair
/a/ => /a:/	81	4	15
/a:/ => /a/	79	11	10
/i/ => /i:/	21	40	39
/i:/ => /i/	15	68	17
/u/ => /u:/	52	20	28
/u:/ => /u/	42	43	15

4. DISCUSSION

4.1. Effet de la quantité

Si nous admettons que la hausse du niveau de l'ambiguïté avec les signaux modifiés est due à la manipulation, alors les signaux "équivoques" sont considérés comme un échec. Ainsi nous n'avons que deux types de résultats "Réussit" et "Échec", alors la confusion faite entre /a/ vs. /a:/ après la neutralisation des différences de durée sera de 80%, ce qui nous permettrait d'avancer l'idée que /a/ et /a:/ auraient une même qualité acoustique réalisée avec deux durées différentes.

L'opposition /i/ vs. /i:/ est très peu sensible à la modification de la durée, les deux voyelles seraient à plus que 80% distinctes en l'absence de l'opposition de leurs durées. /i/ et /i:/ seraient effectivement deux qualités vocaliques bien distinctes. Quant à l'opposition /u/ et /u:/, sa sensibilité au changement de la durée se situe entre les deux autres catégories, avec une légère prédominance de la qualité sur la quantité. Ces observations nous permettent d'attester du fait que les voyelles n'utilisent pas le paramètre de la durée au même degré pour affirmer leurs identités vocaliques.

Notre étude (non publiée) sur la durée de la syllabe en Arabe (Syrie), pour des voyelles longues et brèves a montré que la durée de la syllabe dépendait de la durée de la voyelle, plus la voyelle est longue et plus la syllabe est longue, ($r = 0,89$; $p < 0,001$), ce qui signifie que les autres unités ne réorganisent leurs durées que peu, en fonction de la durée de la voyelle pour maintenir une forme canonique temporelle de la syllabe, par conséquent, la syllabe aurait très peu d'effet sur la perception de la durée de la voyelle qui constitue son noyau. Si, au contraire, la syllabe maintenait la même durée pour longue comme pour la brève, nous ne pouvons nous attendre que à ce qu'il y ait une réorganisation temporelle des consonnes pré et post vocalique, une réorganisation qui servira d'indice pour la durée de la voyelle.

4.2. Effet de la qualité

Dans une étude (non publiée) nous avons réussi à déterminer les distances (en Bark) en termes de F1 et de F2 entre les deux termes de chacune des 3 paires étudiées ici, en fait, les résultats sont en total accord avec ce que nous avons avancé dans le paragraphe précédent. En ce sens que la catégorie dont les deux termes se confondaient le moins lors de la disparition de l'effet de la durée est celle qui contient les voyelles les plus distantes, ce qui signifie aussi que les voyelles de cette catégorie fondent l'opposition de leurs identités sur l'information spectrale. Dans l'ordre, plus les deux termes se confondaient sans l'effet de la durée et moins la distance perceptive est importante. Nos mesures ont montré que la distance perceptive dans l'espace vocalique entre /i/ et /i:/ est 2 fois la distance entre /u/ et /u:/ et autour de 2,4 fois la distance /a/-/a:/. Respectivement 1,44; 0,72 et 0,61 Bark.

5. CONCLUSION ET PERSPECTIVES

Nous avons pu déterminer dans cette étude l'étendue de la contribution de la durée de la voyelle dans le maintien du contraste Brève vs. Longue pour les trois paires des voyelles cardinales de l'Arabe (Syrie). Nous concluons que la durée n'est pour aucune paire le seul paramètre qui permet la distinction des deux termes de ce contraste, les indices acoustiques qui permettent leurs distinctions sont supportés par des informations temporelle et spectrale. Pour la catégorie /a/, l'information temporelle est l'indice acoustique principal qui permet la distinction

Brève vs. Longue. Pour la catégorie /u/, cette information partage son rôle distinctif avec l'information spectrale, quasiment à part égale. Quant à la catégorie /i/ c'est, sans aucun doute, l'information spectrale qui est l'indice acoustique principal.

Cette quête préliminaire, pas finie, est motivée par la description du système phonologique des voyelles de l'Arabe, et ceci en donnant le nombre des qualités vocaliques qui y existent. Ainsi pour notre système considéré comme ayant 3 qualités vocaliques, dédoublées par le trait de la durée, nous proposons que celui-ci a au moins 4 qualités /i:/, /a:/, /u:/ et /i/. Reste à justifier pourquoi il y a autour de 18 % de confusion au sein de la catégorie /i/ en l'absence de l'effet de la durée, et à sonder plus profondément le rôle des durées des unités phonétiques de la syllabe dans la perception de la durée de la voyelle dont elle est le noyau.

REFERENCES

- [Ain72] Ainsworth, W. A. (1972). "Duration as a cue in the recognition of synthetic vowels", *Journal of the Acoustical Society of America*, Vol. 51, pp. 648-651.
- [Ani70] Al-Ani, S. (1970). "Arabic Phonology", *Janua Linguarum, Series Practica*, 61, The Hague/Paris: Mouton.
- [All04] Allatif, O. & Abry, C. (2004). "Adaptabilité des paramètres temporels et spectraux dans l'opposition de quantité vocalique de l'arabe de Mayadin (Syrie)", *Actes des XXVes Journées d'Étude sur la Parole*, pp. 13-16, Fès, Maroc.
- [Hai94] Abou Haidar, L. (1994). "Norme linguistique et variabilité dialectale: analyse formantique du système vocalique de la langue arabe"., *Revue de Phonétique Appliquée*, Vol. 110, pp. 1-15.
- [Bar03] Barkat-Defradas, M., Vasilescu, I. & Pellegrino, F. (2003). "Stratégies perceptuelles et identification automatique des langues: application au continuum dialectal arabe", *Revue Parole*, Vol. 25, pp. 1-44.
- [Got84] Gottfried, T. L. (1984). "Effects of consonant context on the perception of French vowels", *Journal of Phonetics*, Vol. 12, pp. 91-114.
- [Got88] Gottfried, T. L. & Beddor, P. S. (1988). "Perception of temporal and spectral information in French vowels", *Language and Speech*, Vol. 31, pp. 57-75.
- [Mer78] Mermelstein, P. (1987). "On the relation between the vowel and the consonant identification when cued by the same acoustic information", *Perception & Psychophysics*, Vol. 23, pp. 331-336.
- [New02] Newman, D. L. & Verhoeven, J. (2002). "Frequency analysis of Arabic vowels in connected speech", *Antwerp Papers in Linguistics*, Vol. 100, pp. 77-86.