THE ROLE OF VOWEL AND CONSONANT DURATION IN VOWEL LENGTH CATEGORISATION BY DJAMBARRPUYŊU LISTENERS

Kathleen Jepson

The University of Melbourne; ARC Centre of Excellence for the Dynamics of Language
kjepson@student.unimelb.edu.au

ABSTRACT

Australian Indigenous languages are often reported to have compensatory lengthening of consonants after phonemically short vowels. There have, however, been very few perception studies of these languages, and none, to date, focused on vowels. This paper aims to determine whether Djambarrpuyŋu listeners use consonantal lengthening as a cue to vowel length. A forced-choice categorisation task investigated how 19 Djambarrpuyŋu listeners (9M, 10F) use segmental duration in their processing of words minimally distinguished by vowel length. Results show that listeners rely primarily on vowel duration in their categorisation behaviour. However, when the vowel's duration was ambiguous, that is, between the duration ranges of short and long, phonetically long nasals affected the cross-over point of word categorisation. This study supports the primacy of the phonemic vowel length contrast for listeners and also shows that durational context can expand the range of vowel length categories.

Keywords: categorical perception, vowel length, consonant duration, Australian Indigenous languages

1. INTRODUCTION

The categorisation of a phone in speech is the result of a constellation of phonetic cues contributing to the overall perception by the listener [19]. There is often a primary cue which listeners rely on, but there may be other secondary cues which contribute to a listener's perception [20]. In languages that have phonemically contrastive vowel length, the duration of the vowel is cross-linguistically found to be the primary cue used by listeners in categorising vowels as one of the relevant length categories [6,15,16].

Other secondary cues to vowel length, such as spectral information and fundamental frequency (f0), are utilised differently across languages in vowel length perception, as they are in production [16]. Crucially, it is often found that the cues exploited by listeners in perception are those that covary with length category in production. These secondary cues may affect a listener's perception, shifting the boundary at which point a stimulus is judged to belong to the short or long category.

Vowel formant characteristics affect vowel length categorisation in, for example, Thai [1,16,21] and Swedish [6,11]. Whereas, f0 is a secondary cue in vowel length perception for Japanese listeners such that falling f0 was perceived as 'long' earlier in the continuum than stimuli with a level f0 contour [14,16] (c.f. German [16,23]).

In other perception research, the immediate phonetic context has been found to provide cues to phonemic category [20,24]. Considering vowel length perception, an example of a context effect is found in Thai for which longer post-vocalic nasal duration conditioned a higher proportion of short-vowel responses than did nasals with shorter duration [21].

The present paper reports on a preliminary study investigating the perceptual categorisation of vowel length in Djambarrpuyŋu, an Australian Indigenous language of northeast Arnhem Land, spoken by ~4,300 people [2].

1.1. Vowels in Djambarrpuyŋu: Phonetics and phonology

Like many Australian Indigenous languages, Djambarrpuyŋu has a three vowel quality system, triangular in shape, with a length contrast, resulting in six vowels /ɪ, ɪː, ɐ, ɐː, ʊ, ʊː/ [25].[1] Vowel length is contrastive only in the initial syllable of words which is putatively the location of primary stress [25]. This restriction of the contrast to the initial syllable is observed in many Australian languages that have contrastive vowel length [4,9,10].

Vowel phonemes in Australian languages typically display a large degree of allophonic variation, often attributed to coarticulation with neighbouring consonants [10]. Further to this, vowel spectral information is found to covary only minimally with length in many languages, often showing a great deal of overlap between categories. This is the case for Djambarrpuyŋu; long and short vowels have very similar spectral characteristics [13].

Consonantal lengthening is reported to occur following phonemically short vowels in open, word-initial syllables in a number of Australian languages [4]. Similar types of inverse duration relationships are observed in, for example, Icelandic, Norwegian and

Swedish (as discussed in [15]). In some formal phonological accounts of Australian languages, this lengthening is attributed to the necessity for stressed syllables to be bimoraic [3,4]. Phonetically, lengthening may contribute to the encoding of accentual prominence [8]. A claim about the reasons for consonantal lengthening is not proposed here. In Djambarrpuyŋu, some consonants following short vowels have been found to have longer duration than consonants following long vowels [12], and this effect is seen most strongly for consonants following short vowels in open syllables [13].

2. AIMS AND HYPOTHESES

It is not yet known how the phonemic vowel length categories of long and short align with phonetic duration of a vowel in perception for any Australian Indigenous language. Further, it is not known to what extent listeners use the vowel-consonant durational complementarity in categorising vowels as to their length category.

This paper aims to establish how Djambarrpuyŋu listeners use duration of vowels and that of the following consonant in categorising vowel length. Specifically investigated is how listeners categorise a word as being the short- or long-vowel member of a pair of words phonemically distinguished by vowel length alone.

It is hypothesised that Djambarrpuyŋu listeners' perception of vowel length is categorical. Further, vowel duration is predicted to be the primary cue listeners rely on in their categorisation of vowel length. However, it is expected that consonant duration is exploited in the durationally ambiguous region of the vowel continuum—that is, when a vowel is durationally between the ranges of the short and long categories. It is hypothesised that in those cases, a following consonant with shorter duration will result in the vowel being perceived as 'long', whereas when a consonant has longer duration the vowel will be perceived as 'short'. This will be reflected in the listeners' word choice. In this regard, it is hypothesised that consonant duration will influence listener judgement of category boundaries. The perceptual category boundary is hypothesised to occur in the durational region between the vowel length categories observed in production.

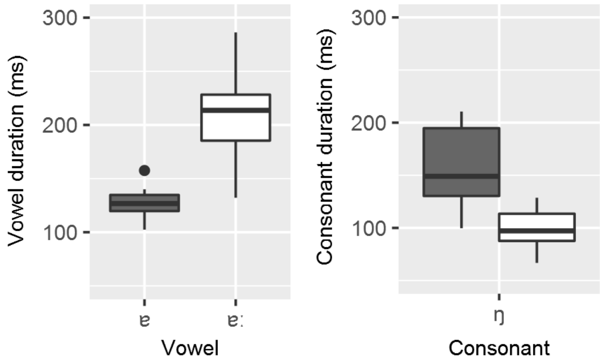
3. METHOD

3.1. Stimuli

The minimal pair *waŋa* /wɐŋɐ/ "talk" and *wäŋa* /wɐːŋɐ/ "home" was selected for investigation in a two-alternative forced-choice categorisation task. These items are comparably frequent in daily speech.

As shown in Figure 1, duration of the intervocalic nasal covaries with vowel length in this pair of words.

Figure 1: Duration values (ms) of vowels /ɐː/ and /ɐ/ and following nasal /ŋ/ in the Djambarrpuyŋu words /wɐŋɐ/ (grey) and /wɐːŋɐ/ (white).



Stimuli were created by manipulating in Praat [7] (using tailored Praat scripts) the duration of the vowel and following nasal in a single token of /wɐŋɐ/, spoken by a Djambarrpuyŋu man, 39 years of age. The duration of the steady-state portions of the first vowel and the following nasal were manipulated along a seven-point continuum which encompassed the durational range observed in the production of those segments in the words /wɐŋɐ/ and /wɐːŋɐ/ in a dataset of seven speakers (plotted in Fig. 1). Steps were evenly spaced along the continua. The durational values of stimuli steps for the vowel and nasal are provided in Table 1. These values represent the duration of the entire segment including the manipulated steady state portion and transitional periods. It is predicted that the category boundary in perception will be between vowel steps 3 and 4 (i.e., 160ms-185ms).

Table 1: Durations (ms) for vowel (V) and nasal (N) steps in stimuli (full factorial design: 7*7 = 49).

Step	V (ms)	N (ms)
1	110	90
2	135	108
3	160	127
4	185	145
5	210	163
6	235	182
7	260	200

3.2. Participants

Nineteen first language Djambarrpuyŋu speakers from Yurrwi (Milingimbi), Northern Territory, Australia, participated in the task (10 women, 9 men; age range = 25–61 years, mean age = 41 years). Participants were compensated for their time.

Hearing loss is not an uncommon issue for people in remote communities in Australia; five participants reported having hearing difficulties in one or both of their ears. The effects that this may have on their speech perception is not investigated here. Participants are representative of their community in terms of age, linguistic repertoire, and hearing.

3.3. Experiment design and procedure

The experiment was constructed and presented in OpenSesame (v.3.1) [17] using a laptop computer, in situ. Participants undertook the task at their homes, either inside or outside. The experiment was full factorial in its design, i.e., included all combinations of manipulated vowel and nasal duration. Stimuli were presented in randomised order in four blocks. Each block consisted of 49 trials (i.e., all vowel step/nasal step combinations). Therefore, each participant completed a total of 196 trials.

Instructions were presented in Djambarrpuyŋu and English onscreen. The researcher discussed the task with each participant. Participants were told the researcher wanted to learn about how Yolŋu (people of northeast Arnhem Land) listened to sounds in words. Participants were instructed to listen carefully and respond thoughtfully but were told that the task was not a test with right answers.

Participants listened to audio stimuli through Philips SHL3060BK over-ear headphones. In each trial, a stimulus was played once. The words /wɐŋɐ/ and /wɐːŋɐ/ were represented visually on screen as two images (Fig. 2). To indicate their choice, participants selected the LEFT SHIFT or RIGHT SHIFT keyboard buttons corresponding to the images on screen. Image and button order remained the same over all trials and blocks. Between blocks, a pause screen invited participants to take a break if they needed. The experiment took approximately 20 minutes to complete, including preliminary discussion and breaks. Background noise and distractions during the experiment existed for all participants. The researcher could pass a trial if the participant was not able to respond. While response times were collected, they are not presented here.

Figure 2: Visual representations of /wɐŋɐ/ "talk" (left) and /wɐːŋɐ/ "home" (right).

3.4. Statistical analysis

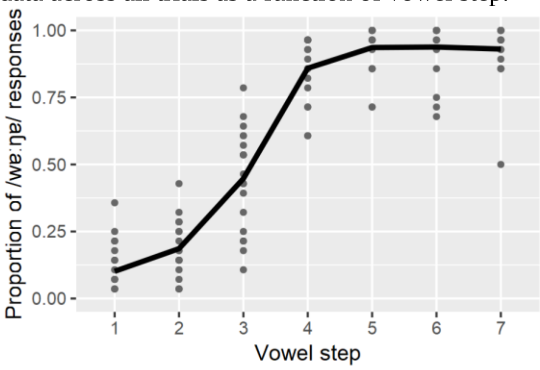
Generalised Linear Mixed Models were used to statistically test the data using the glmer() function from the lme4 package [5] in R [18,22]. Response (/wɐŋɐ/-/wɐːŋɐ/) was the dependent variable, fixed factors were vowel step (seven levels), nasal step (seven levels), block (four levels), as well as an interaction between vowel step and block. The random effects structure included intercepts for participant (19 levels), and by-participant random slopes for the effect of vowel step. The best fitting model was selected using likelihood ratio tests. Selected results of Tukey's post-hoc tests are presented. Results are considered significant if $p < 0.05$.

4. RESULTS

Of the 3,724 total responses, 6 responses were "passed", 1,378 of the responses were /wɐŋɐ/, and 2,340 of the responses were /wɐːŋɐ/. Responses, therefore, are skewed towards /wɐːŋɐ/ (63%).

Figure 3 shows the proportion of /wɐːŋɐ/ responses as a function of vowel step (1 having the shortest duration, 7 having the longest duration). Individual speaker response frequencies are plotted by points. There is an s-curve across the seven steps in the vowel duration continuum. This confirms the categorical perception of these words by vowel length for these Djambarrpuyŋu listeners. Vowel step had a significant effect on response, $\chi^2(24, N = 3718) = 82.47$, $p < 0.001$. Comparisons of steps 1~2, 4~7, 5~6, 5~7 and 6~7 were not significant. All other comparisons were significant; steps 4~5 and 4~6 $p < 0.05$; rest $p < 0.001$. Based on vowel duration, the category boundary (50% cross-over) for these Djambarrpuyŋu listeners is around vowel step 3.

Figure 3: Proportion of /wɐːŋɐ/ responses for all data across all trials as a function of vowel step.



Considering the proportion of /wɐːŋɐ/ responses as a function of nasal duration (Fig. 4), it is clear there is not a categorical shift in perception due to nasal step across the data, though nasal step had a significant effect on response, $\chi^2(6, N = 3718) = 55.47$, $p < 0.001$. There is a slight decrease in /wɐːŋɐ/

responses when the nasal duration is at the longest point in the continuum.

Figure 4: Proportion of /wɐːŋɐ/ responses for all data across all trials as a function of nasal step.

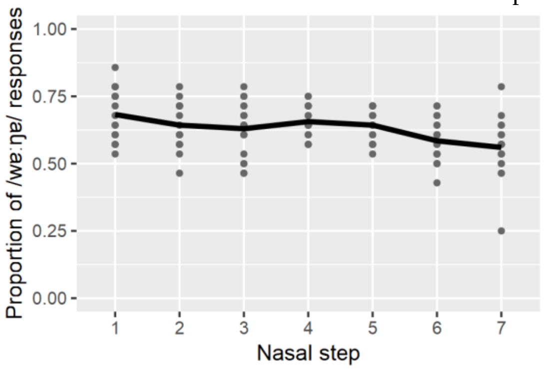


Figure 5 presents /wɐːŋɐ/ responses as a function of nasal step for stimuli containing a vowel at step 1, step 3 and step 7. Comparing the responses for stimuli which include vowels at the extremes of the vowel step continuum, listeners reliably selected /wɐŋɐ/ when presented with a stimulus containing a step 1 vowel, and conversely /wɐːŋɐ/ when presented with a stimulus containing a step 7 vowel, irrespective of nasal step. Step 3 vowel data, as anticipated, represent the "ambiguous" region in the vowel duration continuum. In vowel step 3 data, nasal step significantly influenced listeners' categorisation, $\chi^2(6, N = 531) = 27.98$, $p < 0.001$. Post-hoc tests showed nasal steps 6 and 7 conditioned fewer /wɐːŋɐ/ responses; 1~7 $p < 0.05$, 1~6, 3~6, 3~7 $p < 0.01$.

Figure 5: Proportion of /wɐːŋɐ/ responses for vowel step 1 data (full line), vowel step 3 data (dashed line), and vowel step 7 data (dotted line) as a function of nasal step.

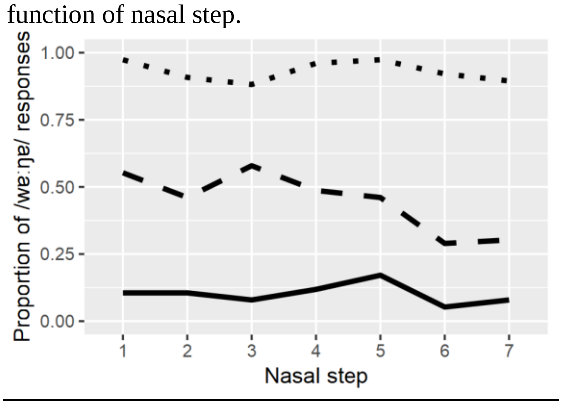
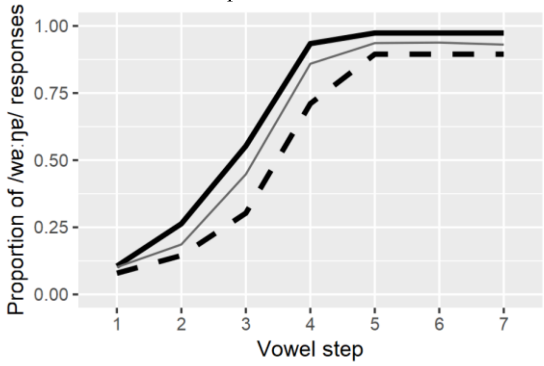


Figure 6 presents /wɐːŋɐ/ responses as a function of vowel step for stimuli containing a step 1 nasal, step 7 nasal, and across all data. Compared with all data, stimuli including step 1 nasals resulted in the category boundary occurring ~0.3 steps earlier than across all the data, whereas stimuli including step 7 nasals resulted in the category boundary occurring ~0.3 steps later.

Figure 6: Proportion of /wɐːŋɐ/ responses for nasal step 1 data (full line) and nasal step 7 data (dashed line) as well as across all data (thin line) as a function of vowel step.



5. SUMMARY AND DISCUSSION

The aim of this study was to explore how vowel duration and that of the following consonant are used in the categorisation of Djambarrpuyŋu words which contrast minimally by vowel length. Results show that the categorisation of words as /wɐŋɐ/ or /wɐːŋɐ/ by Djambarrpuyŋu listeners was based primarily on the duration of the vowel. Overall, the category boundary in perception appears to be aligned with the production categories for these vowels [13]. Nasal duration influenced the categorisation of stimuli when the vowel was between the durational regions of the short and long categories. For those stimuli, nasals with longer duration conditioned more /wɐŋɐ/ (short vowel) responses. This finding concords with previous research for which additional features or secondary cues (i.e., nasal duration in the present study), "are largely ignored unless the primary feature is impaired in some way" [24]. These results support the analysis of vowel length in Djambarrpuyŋu being phonological, while also showing that listeners' perception is influenced by the duration of the following consonant.

To better understand the extent to which the post-vocalic consonant duration context influences listener perception, a further experiment could be undertaken using stimuli created by manipulating a token of /wɐːŋɐ/. Other questions raised by this exploratory experiment include: what other cues might listeners be attuning to in their perception? It is hypothesised that the alignment of f0 contour may play an important role as it appears to in production [13]. Do other vowel-consonant sequences behave similarly, or is this behaviour limited to a subset of vowels and consonants? It is clear that further perception research will be fruitful in ongoing descriptive work on the phonetics and phonology of the Djambarrpuyŋu language.

6. ACKNOWLEDGEMENTS

This research was conducted with support from the ARC Centre of Excellence for the Dynamics of Language (Project ID: CE140100041). Many thanks to Albena Buyanggirr for translating the instructions for the task from English to Djambarrpuyŋu. I thank all Yolŋu who participated in this study. Thanks to Jonathan Harrington for encouraging me to conduct a perception experiment. Thank you to my supervisors Janet Fletcher, Hywel Stoakes and Ruth Singer, as well as my colleagues in the Phonetics Lab, and three anonymous reviewers for helpful comments and suggestions.

7. REFERENCES

[1] Abramson, A. S., Ren, N. 1990. Distinctive vowel length: Duration vs. spectrum in Thai. *Haskins Laboratories Status Report on Speech Research*. 101/102, 256–268.

[2] Australian Bureau of Statistics. 2016. *Australia, census 2016: Language spoken at home by sex (SA2+)*. stat.data.abs.gov.au/Index.aspx?DataSetCode=ABS_C16_T09_SA

[3] Baker, B. 2008. *Word structure in Ngalakgan*. Stanford, CA: CSLI.

[4] Baker, B. 2014. Word structure in Australian languages. In: Koch H., Nordlinger R. (eds), *The languages and linguistics of Australia*. Berlin: Mouton de Gruyter, 139–213.

[5] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. 67, 1–48.

[6] Behne, D., Arai, T., Czigler, P., Sullivan, K. 1999. Vowel duration and spectra as perceptual cues to vowel quantity: A comparison of Japanese and Swedish. *Proc. 14th ICPhS* San Francisco, 857–860.

[7] Boersma, P., Weenink, D. 2017. *Praat: Doing phonetics by computer*. Version 5.3.

[8] Butcher, A. 2006. Australian Aboriginal languages: Consonant-salient phonologies and the "Place-of-Articulation Imperative". In: Harrington J., Tabain M. (eds), *Speech Production*. Hoboken: Taylor and Francis, 187–210.

[9] Dixon, R. M. W. 2002. *Australian languages: Their nature and development*. Cambridge: CUP.

[10] Fletcher, J., Butcher, A. 2014. Sound patterns of Australian languages. In: Koch H., Nordlinger R. (eds), *The languages and linguistics of Australia*. Berlin: Mouton de Gruyter, 91–138.

[11] Hadding-Koch, K., Abramson, A. S. 1964. Duration versus spectrum in Swedish vowels: Some perceptual experiments. *Studia Linguistica*. 18, 94–107.

[12] Jepson, K., Stoakes, H. 2015. Vowel duration and consonant lengthening in Djambarrpuyngu. *Proc. 18th ICPhS* Glasgow.

[13] Jepson, K. *Prosody, prominence and segments in Djambarrpuyŋu.* PhD thesis, Univeristy of Melbourne. Forthcoming.

[14] Kinoshita, K., Behne, D., Arai, T. 2002. Duration and f0 as perceptual cues to Japanese vowel quantity. *Proc. 7th INTERSPEECH* Denver, 757–760.

[15] Lehnert-LeHouillier, H. 2007. *The perception of vowel quantity: A cross-linguistic investigation*. PhD thesis, State University of New York at Buffalo.

[16] Lehnert-LeHouillier, H. 2010. A cross-linguistic investigation of cues to vowel length perception. *J. Phon.* 38, 472–482.

[17] Mathôt, S., Schreij, D., Theeuwes, J. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*. 44, 314–324.

[18] R Core Team. 2017. *R: A language and environment for statistical computing*. Version 3.4.2. The R Foundation for Statistical Computing, Vienna.

[19] Raphael, L. J. 2005. Acoustic cues to the perception of segmental phonemes. In: Pisoni D. B., Remez R. E. (eds), *The handbook of speech perception*. Oxford: Blackwell, 182–206.

[20] Repp, B. H. 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*. 92, 81–110.

[21] Roengpitya, R. 2001. *A study of vowels, diphthongs, and tones in Thai*. PhD thesis, UC Berkeley.

[22] RStudio Team. 2016. *RStudio: Integrated development for R.* Version 1.1.383. RStudio, Inc., Boston.

[23] van Dommelen, W. A. 1993. Does dynamic f0 increase perceived duration? New light on an old issue. *J. Phon*. 21, 367–386.

[24] Whalen, D. H., Abramson, A. S., Lisker, L., Mody, M. 1993. F0 gives voicing information even with unambiguous voice onset times. *JASA*. 93, 2152–2159.

[25] Wilkinson, M. 2012. *Djambarrpuyŋu: A Yolŋu variety of northern Australia*. Munich: Lincom Europa.

---

[1] Accompanying these vowels there are 25 phonemic consonants; six places of articulation for stops (two series) and nasals, as well as two glides, two rhotics, two lateral approximants, and a glottal stop.