Learning vowel categories from maternal speech in Gurindji Kriol

To appear in *Language Learning, 62*, accepted 16 April 2011

Caroline Jones¹, Felicity Meakins², Shujau Muawiyath¹

¹University of Wollongong, Australia
²University of Queensland, Australia
Abstract

Distributional learning (Anderson, Morgan, & White, 2003; Maye, Werker, & Gerken, 2002) is a proposal for how infants might learn early speech sound categories from acoustic input before they know many words. When categories in the input differ greatly in relative frequency, and overlap in acoustic space, research in bilingual development (Bosch & Sebastián-Gallés, 2003; Sebastián-Gallés & Bosch, 2009) suggests that this affects the course of development. In the present study we describe the nature and extent of vowel variation in nearly 900 vowel tokens in maternal speech in Gurindji Kriol, a mixed language of northern Australia, which like bilingual input has differences in the relative frequency of adjacent, overlapping vowel categories. In Analysis 1, we provide the first systematic account of vowel variation and phone frequency in maternal speech in Gurindji Kriol. In Analysis 2, cluster analysis was applied to the vowel formant and duration data, to see what categories might emerge from acoustic data alone. The results suggest that, were infants to base their initial vowel categories solely on the clusters emerging in acoustic space, they might likely set up relatively few vowel categories. We discuss implications for how infants may learn Gurindji Kriol and for distributional learning.
How do infants learn vowel and consonant categories in their native language, before they have learned few words, and what is the path of this development? Recent research has made progress in this area by exploring how far it is possible to take the distributional learning hypothesis (e.g., Anderson, Morgan, & White, 2003; Maye, Werker, & Gerken, 2002): that vowel and consonant categories are initially built in memory in response to the experience of statistical distributions of sounds in the language(s) to which the infant is exposed. A key aspect of such experience is the shape of the distribution or distributions (Maye, Werker, & Gerken, 2002; Maye, Weiss, & Aslin, 2008) and the extent of overlap between distributions of sounds that are adjacent in acoustic space (Bosch & Sebastián-Gallés, 2003). More recently, some of the research investigating category learning by infants growing up in bilingual environments has been interpreted (by Burns, Yoshida, Hill, and Werker, 2007; Sebastián-Gallés and Bosch, 2009; Sundara, Polka, & Molnar, 2008) as supporting a key role for the relative sizes of distributions of sounds, being due to the relative frequency of different sounds in the language experience of the infants, summed over both native languages. In the present research we consider the implications for distributional learning from a type of maternal speech data that like bilingual input is characterised by overlapping distributions and substantial differences between the frequencies of individual speech sounds -- vowels in a recently developed Australian mixed language, Gurindji Kriol.

Background to the study

In this section, we first review the distributional learning hypothesis, then the relevant research into category learning from bilingual exposure, and then explain why Gurindji Kriol maternal speech research is an interesting case to consider for the distributional learning hypothesis. Following the literature review, we describe the
nature and extent of vowel variation in Gurindji Kriol (Analysis 1), and consider how
distributional learning would fare with this input (in Analysis 2).

The distributional learning hypothesis

The distributional learning hypothesis has become a popular idea about how
infants might initially form phonetic categories. In essence, the idea is that infants
might use statistical patterns of experience to form initial phonetic categories from
experiences that clump together in psychophysical space, typically involving auditory
parameters of speech sounds (e.g. F1, F2 for vowels, or voice onset time for
consonant voicing). A main attraction of this view is that it might explain how infants
develop perceptual sensitivity to contrasts in their native language within the first year
of life, at an age when their experience and knowledge of words would seem too
limited to allow them to use evidence from minimal pairs as a guide to vowel and
consonant categories.

There is now some evidence to support the plausibility of this idea, from
analyses of adult production data, experimental studies with infants and from lab-
based studies of infant-directed speech. In adult production data, there is evidence of
language-specific distributions of acoustic cues. For example, vowel contrasts in
maternal speech (Werker, Pons, Dietrich, Kajikawa, Fais, & Amano, 2007) and in
adult speech (Hillenbrand, Getty, Clark, & Wheeler, 1995) are supported by patterns
of acoustic variation (in e.g. duration, formant frequencies) which are language-
specific. There are also acoustic data from consonants which support the idea that
phonemes correspond to peaks in the acoustic distribution, such as finding two peaks
in the distribution of voice onset time in stop tokens in languages like English which
have a voiced-voiceless contrast (Lisker & Abramson, 1964; Lotto, Sato, & Diehl,
In an experimental study, Maye, Werker, and Gerken (2002) exposed infants to either a unimodal (one-peak) or bimodal (two-peak) distribution of sounds varying in voice onset time (VOT). Infants who had been exposed to the bimodal distribution were able to categorize new tokens into two VOT categories corresponding to the two endpoints. But infants who had been exposed only to the unimodal distribution were not able to categorize the endpoints as accurately.

There is also evidence from classification-based analysis of infant directed speech (Vallabha et al., 2007, using read data from nonce words in infant directed recordings from Werker et al., 2007). In that study, the algorithm was able to classify vowel phonemes (English /ɪ, i, ɛ, e/ and Japanese /i, i:, e, e:/) with considerable accuracy, from single measures of F1 and F2 and from duration of steady state. As noted in a review by Swingley (2009), the speech sample analysed was elicited under lab conditions and may be different from naturalistic infant directed speech. For example, lab-recorded infant directed speech may differ from natural infant directed speech in the extent of articulation of segments, and in speech rate and prosody. The contextual variation in lab-recorded infant directed speech is also limited since the vowels tend to be in phonetically controlled words.

In recognition of these limitations, Swingley (2009) briefly presents some vowel data drawn from a naturalistic sample (700 vowels from one mother, recordings from the Brent & Siskind, 2001, database), and discusses the apparent clustering of the vowels in plots of F1, F2 and vowel duration (coded by actual phoneme). Swingley inspects the plots to argue that if distributional learning is construed essentially as context-free cluster analysis, then in the example examined,
this learning would seem unlikely to find distinct clusters corresponding to vowel phonemes (at least when the data are F1, F2 and vowel duration, one datapoint per vowel). This state of affairs appears consistent with the results of a previous study which aimed in part to uncover adult vowels using acoustic data (Hillenbrand et al., 1995). In that paper, non-IDS adult vowels (in controlled phonetic frames) were not as well separated by discriminant analysis when the vowel measures were taken at one timepoint per vowel, as when spectral change was measured.

The import of these kind of results for distributional learning hypothesis is still unclear, and depends on the version of the hypothesis which is adopted. As Swingley (2009) argues and notes as a suggestion by Davis and Lindblom (2001), it is not out of the question that distributional learning would succeed in finding categories if enough parameters were included. For example, the problem of establishing vowel categories might be overstated by results deriving from analyses which focus on static acoustic data about vowels. There is a great deal of previous research showing that adult listeners use contextual information to perceive phonetic categories, and some research showing that children use contextual information, too. In studies of context effects (Repp, 1980), listeners' perceptions of a sound are systematically affected by phonetic context. For example, adults' use of F3 at vowel onset in perception of prevocalic /s/ vs /ʃ/ depends on the following vowel. In a study by Mann and Repp (1980), adults used F3 information more when the vowel was /u/ than when it was /a/, probably since /u/ intrinsically has a higher F3. Child listeners also display this context effect in their perception, although the context effect is less evident for children under 7 years, as shown in the results of Nittrouer and Miller (1997). Research into vowel perception by children, using specially constructed stimuli, has
so far found conflicting results concerning the extent to which children may use contextual / dynamic information (Nittrouer, 2007; Sussman, 2001).

The parameters which would simplify the problem of learning categories from variable input are not limited to acoustic variables, even though these have traditionally been the focus of speech perception research. It seems plausible that visual cue (e.g. facial gestures) and covariance related to higher levels of linguistic knowledge (such as words) and social knowledge (socioindexical variation such as gender and accents) could be used by infants. Among the latter sources, Swingley (2009) proposes that early lexical knowledge used in word identification could help point the infant towards early phonetic categories, in addition to the evidence from the phonetic level.

In the present study, we analyse a sample of natural infant-directed speech which coincidentally resembles the dataset discussed by Swingley (2009). We present the results of cluster analyses on F1, F2 and vowel duration data from 879 monophthongal vowel tokens from the speech of one mother. The data include schwa vowel tokens. These naturalistic data are a valuable testing ground for the distributional learning hypothesis. Since the data come from a mixed language, they also provide an opportunity for exploring further some recent ideas about the effects of category frequency in the input, stemming from the bilingual literature.

Learning speech categories from bilingual input

The recent increase in research into speech category learning among infants learning in bilingual contexts has sparked discussions in the literature about a possible role for frequency, that is, the relative numbers of tokens of phonemes in the input. Bosch and Sebastián-Gallés (2003) found U-shaped development during the first year of life for infants from Catalan-Spanish bilingual homes, in discriminating Catalan /ɛ/
and /e/. The development is termed "U-shaped" since the bilingual infants seem to perceive the contrast early (at 4 months), show a drop in success (at 8 months), and then again show perception of the contrast late in the first year (at 12 months). In the acoustic input, tokens of Catalan /ɛ/ and /e/ tend to cluster in the front vowel space, below and above Spanish /e/, respectively. The three distributions overlap in F1/F2, likely presenting infant learners with something very much like a single-peak distribution in the front mid vowel space, as Bosch and Sebastián-Gallés argue (Bosch & Sebastián-Gallés, 2003; Sebastián-Gallés & Bosch, 2009). What also seems to be a contributing factor is that Spanish /e/ is relatively much more frequent (at approximately 25% of vowel tokens in Spanish) than Catalan /ɛ/ or /e/ (which together account for less than 9% of vowel tokens in Catalan), as those authors argue.

The suggestion that frequency of tokens of a given phoneme affects infants' learning is consistent with the results of other studies with infants from bilingual homes. These other studies have found that bilingual infants keep pace with monolingual peers in perceptual reorganisation in the first year of life. Burns, Yoshida, Hill, and Werker (2007) found that for stop voicing (voice onset time manipulations), infants from French-English bilingual homes discriminated both English and French voicing contrasts at 10-12 months on the same time frame as monolingual infants. Burns, Yoshida, Hill, and Werker (2007) noted in their discussion that the contrast they investigated is highly frequent in the ambient language, which as they say may account for the development of sensitivity to this contrast at 10-12 months. In a similar line of reasoning, the results of Sundara, Polka, and Molnar (2008) were ascribed to the high frequency of /d/ in French and English: it seems likely that infants from bilingual homes discriminated typically dental French /d/ tokens from typically alveolar English /d/ tokens because they are highly frequent
in the input, even though they come from overlapping categories in acoustic space. Frequency thus interacts with overlap, on this account.

More recent research by Bosch and Sebastián-Gallés (2009) with Catalan-Spanish infants has found evidence for U-shaped development with two other contrasts: /o/-/u/ and /s/-/z/. The interpretation of these results appears much less straightforward, since these do not appear to present a unimodal distribution nor a large frequency difference. Bosch and Sebastián-Gallés (2009) note that /o/-/u/ is, however, relatively overlapping in acoustic space. The extent to which these results can be explained through the factors currently seen as central to distributional learning - overlap and frequency - remains to be seen.

*Vowels in Gurindji Kriol*

Like the case of Catalan-Spanish bilingual infants learning front mid vowels, another case that involves overlap and frequency disparities is vowel input in Gurindji Kriol, a modern Australian Aboriginal language spoken in northern Australia. Gurindji Kriol offers category overlaps plus what seem to be large frequency differences between vowel categories due to its emergence as a ‘mixed language’.

Mixed languages are the result of the fusion of two identifiable source languages, normally in situations of community bilingualism (Matras & Bakker, 2003). The most common type of mixed language fuses the grammar of one language with the lexicon of another. Most rare are the mixtures where both source languages contribute grammatical and lexical material. The most well-known example of this type of mixed language is Michif, which is spoken by the descendents of French Canadian fur traders and Cree women (Bakker, 1997).

Gurindji Kriol is also an example of this rare type; it is a systematic fusion of the lexicon and grammar of Gurindji (a traditional Australian Aboriginal language) and
Kriol (an English lexifier creole spoken in northern Australia). It is thought that Gurindji Kriol developed as children in the late 1970s and 1980s learned a language from listening to Gurindji code-switched with Kriol, as was common among adults at that time (McConvell & Meakins, 2005). Gurindji Kriol is no longer itself a case of code-switching; adults independently produce very similar grammatical constructions in picture description tasks, for example (Meakins, to appear).

Gurindji Kriol has a grammatical system where the nominal grammar including case-marking and derivational morphology is derived from Gurindji and the verb phrase structure including tense-mood-aspect markers comes from Kriol. Lexically it is also very mixed with both source languages contributing nouns and verbs. In the broader corpus of Gurindji Kriol recordings made by Felicity Meakins, approximately two-thirds of the words are derived from Kriol, and around one-third from Gurindji. In terms of morpheme types, 31% (526/1722) are Gurindji-derived and 69% (1187/1722) are Kriol-derived. In terms of morpheme tokens, 30% (55369/184150) are Gurindji-derived and 70% (128781/184150) are Kriol-derived. A very small set of words derive from another traditional Australian Aboriginal language (Warlpiri) or are derivationally mixed. See Meakins (2008, 2009, 2010, 2011) for more details. An example of a typical sentence in Gurindji Kriol that illustrates the nature of this language is given in illustration (1), in (a) in orthography, glossed in (b), and translated in (c). Gurindji-derived elements are italicised and Kriol-derived elements are represented in plain font:

(1)

(a) det warlaku-ngku bin bait-im det marluka fut-ta

(b) the dog-ergative past bite-transitive the old.man foot-on

(c) The dog bit the old man on the foot.
The vowel system of Gurindji Kriol is the focus of the current study. Gurindji Kriol sound patterns reflect the mix of words in the lexicon. Words in Gurindji Kriol are derived from Gurindji, which traditionally has three vowel phonemes /ɪ, ɐ, ʊ/, and from Kriol, which in the local region (the town of Katherine, Northern Territory, and environs) has been regarded as having a five vowel system /ɪ, ɛ, ɐ, ɔ, ʊ/. In the synchronic analysis of Gurindji Kriol phonology that we begin in the current study, the monophthong vowels of this language appear to comprise five phonemes /ɪ, ɛ, ɐ, ɔ, ʊ/. Examples of minimal and near-minimal pairs indicating contrasts among vowels /ɪ, ɛ, ɐ, ɔ, ʊ/ are shown in illustration (2), with the five contrast comparisons in (a) for /ɪ/, plus remaining contrasts illustrated for (b) /ʊ/, (c) /ɐ/, (d) /ɛ/ and /ɔ/. We include the orthographic form for words derived from Gurindji. Words without this annotation are derived from Kriol words whose English origins are clear as they are cognate with the gloss. In square brackets after the gloss, we provide speaker initials with permission, the file name, and the time code for the illustrative example:

(2)

(a) /ɪ/

ɲɪɪ ‘nose’ [CA: FM032.B 212551_214687] (from Gurindji jitji ‘nose’)

ɲʊɲ ‘a sore’ [SS: FM029.B.AUD 621697_624391] (from Gurindji juj ‘a sore’)

stɪk ‘stick’ [SS: FM029.B.AUD 216323_218738]

stɛk ‘stuck’ [SS: FM053.A 193713_196686]

tɪn ‘tin’ [CA: FM041.C 666698_669856]

tɛn ‘ten’ [SS: FM052.B 222081_225053]
(b) /ʊ/

gʊd ‘good’ [SS: FM005.B 134888_138835]
gʊt ‘with’ [SS: FM028.A 719609_722627]

lʊŋk ‘look’ [AR: FM037.B 77246_78290]
lʊk ‘locked’ [SS: FM032.A 841591_845352]

pʊrm ‘put’ [SS: FM005.B 78457_82033]
pɛt ‘pet’ [SS: FM053 818610_820329]

(c) /ɛ/

bɛktɛ ‘bucket’ [SS: FM040.B 81353_84047]
bɛk ‘back’ [SS: FM005.B 578245_584886]

dɛkdeɛk ‘duck’ [CA: FM032.B 273974_277503]
dɔkte ‘doctor’ [SS: FM036.A 1221851_1222896]

(d) /ɛ/ and /ɔ/

nɛk ‘neck’ [AR: FM035.B 1466235_1468162]
nɔk ‘knock’ [CA: FM028.A 1880574_1882710]
There is considerable vowel variation in Gurindji Kriol within words, which is included in the linguistic description in Analysis 1 below. Although the minimal and near-minimal pairs just presented are words in which there is consistency in the vowel quality across different tokens, other words display considerable variation in vowel quality involving words containing [ɔ] in particular.

There are also frequency differences between the vowels in Gurindji Kriol. In terms of numbers of tokens, [ɔ] has much lower frequency than the vowels [ɪ, ɐ, ɛ, ʊ]. Gurindji Kriol thus presents a relevant kind of language input for considering the possible impact of frequency disparities on acquisition.

As well as contributing to research on distributional learning in first language acquisition, the present research also adds to our understanding of mixed languages and their acquisition. Gurindji Kriol is a language of young people; it is the everyday language of adults under about 37 years of age in traditionally Gurindji speaking communities of northern Australia (Meakins, 2008). Today, children in traditionally Gurindji communities are learning Gurindji Kriol as their first language, in a context where Gurindji Kriol is the everyday language variety spoken between adults and children. The children are also growing up in a multilingual environment. They are exposed to traditional Gurindji heavily code-switched with Kriol in the speech of the very oldest people in the community (though this exposure depends partly on family and/or household membership), but are not learning to speak fluent traditional Gurindji. Children are taught in English when they enter the primary school in the community at around age six. In the community, English is used in official domains, at school, clinic, and government offices. In home life, English is only used in fun, when imitating nurses or teachers, for example.
Little systematic research has previously examined the phonetics and phonology of Gurindji Kriol, or how children learn mixed languages in general (although see O’Shanessy, 2008, 2009, to appear, for studies of the acquisition of ergative marking in Light Warlpiri, another Australian Aboriginal language which features a mixture of words and grammar from Kriol and from a traditional Australian Aboriginal language, in this case Warlpiri). There has been much more research and debate on the typology and sociolinguistic status of mixed languages (Golovko, 2003; Matras & Bakker, 2003; Muysken, 1997; van Gijn, 2009) than on the acquisition of these languages by children.

This paper presents in Analysis 1 below the first systematic descriptive account of vowel variation in Gurindji Kriol. In this we analyse a sample of maternal speech in Gurindji Kriol using counts and patterns in the phonetic transcription to describe the vowel variation in Gurindji Kriol and to illustrate how some vowels are more frequent than others in Gurindji Kriol.

In Analysis 2, we consider for the first time in a mixed language the possible impact of exposure to frequency differences among vowels in the input, on the formation of vowel categories in young children. To do this we use cluster analyses of the same maternal speech sample to explore what kinds of vowel categories an infant learning Gurindji Kriol might initially set up, based on certain distributional data alone. On the question of what categories these are, we note that distributional learning is typically presented in research studies as a mechanism for learning phonemes (and in vowel studies, the input data are typically limited to full vowels). Distributional learning might also (or instead) be a way of learning phones, perhaps positional phones (which are early categories in the model of Pierrehumbert, 2003, for
example). We do not assume that the children's first categories are necessarily phonemes, and our analyses do not rely on this assumption.

The present study

Analysis 1: Transcription-based account of vowel categories

The sample of maternal speech used in this study comes from a large database of recordings made by Felicity Meakins. That project studied the speech of young Aboriginal people in northern Australia in everyday informal conversation with their families, including young children. The current sample comprises the speech of one mother speaking often to her young son (aged 2;6;0), but also to other adult members of the family in a group context in which he could overhear. This maternal speech sample reflects what is probably a common situation where infants and young children learn from experiencing both infant- and adult-directed speech (Oshima-Takane, 1988). In the sample analysed here, the majority of utterances are directed to the child. It should be noted that the speech rate in this sample is not strikingly slow or prosodically exaggerated, unlike typical descriptions of infant-directed speech in English, for example.

Method

The maternal speech sample was transcribed into International Phonetic Alphabet (IPA) symbols by the first author using Phon (Rose, Hedlund, Byrne, Wareham, & MacWhinney, 2007). An effort was made to transcribe the vowels narrowly into as many different IPA vowel phones as the native English speaking transcriber could reliably resolve. As the recordings are of conversational speech in which duration varies greatly by prosodic context, and where the speech is quite fast, variation in
vowel duration is not reliably transcribed in the sample and so was not included in this analysis. We present data here on phones transcribed as monophthongs only, although there are diphthongs in the sample.

During the process of transcription, checks on the transcription were made manually with reference to the vowel acoustics using Praat 5.0.11 (Boersma & Weenink, 2007). Vowels were retranscribed where there appeared to be gross differences between the symbols and the acoustic values (using reference data from Harrington, Cox, & Evans, 1997, a study of Australian English).

During a second pass of checking and editing, a systematic effort was made to ensure that the IPA transcription was an accurate and consistent representation of the vowel variation in acoustic terms. Vowel intervals were labelled in Praat, using as landmarks the first pitch period of the vowel (for vowel onset) and the termination of high frequency energy in the spectrogram or last regular pitch period (for vowel offset). This labelling task was shared by first and third authors. The first and third authors checked each others’ labelling and reached consensus on discrepant labelling. Praat scripts calculated F1 and F2 for each vowel token in Hertz, at the temporal midpoint of the vowel. F1 corresponds to vowel height, and F2 corresponds to vowel frontness-backness. The Praat analysis used the Burg algorithm, set to look for 5 formants with a maximum formant value of 5500 Hz (appropriate for an adult female speaker), with a 25 ms analysis window (for an effective window length of 50 ms), applied to the .wav file with a 44.1 kHz sampling rate.

The transcription and acoustic data were screened for potential inaccuracies due to transcriber error and/or autoformant tracking errors by examining apparent outliers among the acoustic values for tokens with particular symbols. Transcription was rechecked and formants examined by hand for tokens whose acoustic values were
grossly inconsistent with the IPA symbol (e.g., an F2 value over 2000 Hz or under 1200 Hz for a phone transcribed as [ɐ]). In this way 45 of the 894 vowel tokens were edited: 27 were edited for IPA symbol used, 16 were edited for acoustic values, 2 for both symbol and acoustic values. A further 15 tokens whose measurement had yielded odd values were removed from the analysis because they could not be acoustically analysed (for various reasons including background noise, speaker overlap, and distortion). Figure 1 provides an overall picture of the F1 and F2 formant values for all the transcribed vowel phone tokens, following checking and editing. In Figure 1 (and subsequent vowel plots in this paper), the vertical axis shows F1 in reverse order, so that high vowels are high in the plot and low vowels are low in the plot. Figure 1 also shows F2, where vowels on the left side are front vowels and vowels on the right side are back vowels. Each IPA symbol in Figure 1 marks an individual vowel token.

---

Table 1 shows the relative frequency of different phones in the sample. The most common phones are [ɪ] and [ɐ] which account for 24.5 and 27.8% of the vowel phones, respectively. The phones [ʊ], [ɛ], and [ə] are considerably less common; these each account for 11-13% of vowel phones. The phone [æ] is very uncommon, at 3.8%, and so are [ʊ] (1.9%), and [ɔ] (2.4%). Other extremely uncommon vowel phones are [e], [o], [ɜ]. (It should be also noted that the ranges of [e] and [o] overlap with [ɛ] and [ɔ], respectively, which makes the use of the symbols [e] and [o] seem less reliable.)

---
Using the transcription, a descriptive analysis was conducted as follows. The description of vowel variation was constructed manually, by observing the possible variant vowel productions in instances of the same word, by organising the words containing each vowel phone by source language phoneme, and by observing patterns of consonant-vowel co-occurrence using electronic searches in Phon.

Results

The transcription data appear consistent with five monophthong vowel phonemes /ɪ, ɛ, ə, ɔ, ʊ/. Each vowel has a range of variation in production within particular words and depending on phonetic context; this variation is described below. Many of the patterns of within-word variation which depend on phonetic context (i.e., surrounding consonants) seem similar to patterns in traditional Gurindji.

Contextual variation. There is much within-word variation between [ɪ], [ʊ] and [ʉ] where the vowel is in the environment of a preceding or following labial or labiovelar consonant. Examples in (3) illustrate this variation for some word tokens:

(3)

‘wait’:

lwed [SS: FM029B 008:08.812 008:11.366]

luwed [SS: FM029B 000:09.317 000:11.151]

(past tense):

brn [SS: FM029B 009:30.807 009:32.943]

‘which way’:

\textit{wiz}c [SS: FM029B 001:02.832 001.04.295]


There is within-word variation between [ʊ] [ɪ] or [ɨ], following or preceding an alveopalatal consonant, as shown in (4):

(4)

‘you’/‘you’-Possessive/‘you’-Ergative/ (Gurindji-derived form):

ɲʊndʊ [SS: FM029B 001:09.277 001:11.668]

ɲʊnʊŋ [SS: FM029B 002:01.420 002:03.695]

ɲɪntʊŋgu [SS: FM029B 002:08.076 002:11.096]

Words which in slower or more deliberate speech can contain an [ɐ] pronunciation are pronounced with [æ] or [ɛ] in the immediate environment of an alveopalatal consonant, illustrated in (5):

(5)

‘throw’:

ʧek [SS: FM029B 003:22.365 003:24.965]

ʧæŋm [SS: FM029B 002:42.928 002:44.832]

ʧekɪm [SS: FM029B 002:37.241 002:40.724]
‘small, little’:

\text{ji\textipa{ bukejw}e\textipa{ [SS: FM029B 011:49.343 012:00.907]}}

\text{ja\textipa{ buke} [SS: FM029B 002:44.786 002:47.015]}

\text{je\textipa{ bukejw}e\textipa{ [SS: FM029B 007:55.039 007:58.058]}}

Within words, there is variation between [ɛ] and [ʊ] or [ɔ], in the environment of a following [p, ŋ, or ŋ], as can be seen in the words in (6):

(6)

‘swim’:

\text{dɛ\textipa{ ruk}e\textipa{ [SS: FM029B 000:20.992 000:22.362]}}

\text{dɛ\textipa{ ruk}e\textipa{ [SS: FM029B 000:40.570 000:43.333]}}

‘big’:

\text{jo\textipa{ nke\textipa{ ni\textipa{ w}e\textipa{ [SS: FM029B 007:42.719 007:45.180]}}}

\text{jo\textipa{ nke\textipa{ ni\textipa{ w}e\textipa{ [SS: FM029B 009:37.377 009:45.411]}}}

\textit{Schwa.} In unstressed position and particularly in fast conversational speech, schwa is a common variant of monophthong vowels. Illustrations are given in (7):

(7)

‘look’:

\text{l\textipa{ uk} [SS: FM029B 001:15.346 001:17.760]}
lsk [SS: FM029B 000:22.548 000:24.986]

(discourse particle):

ne [SS: FM029B 001:25.437 001:29.431]

no [SS: FM029B 000:20.992 000:22.362]

*Possible free variation.* Within-word variation between [ʊ] or [ɔ] is found in some Kriol-derived words, in apparent free variation, as in (8):

(8)

'go':

gon [SS: FM029B 001:02.832 001:04.295]

gun [SS: FM029B 007:16.240 007:17.819]

'come':

kom [SS: FM029B 001:07.419 001:09.346]

kem [SS: FM029B 07:50.753 007:55.165]

Some words are pronounced on different occasions with [ɛ], [æ] or [ε]. This variation is found in some very common Kriol-derived words, shown in (9):

(9)

'and':

‘have’:

æbm [SS: FM029B 004:25.168 004:29.348]

æb [SS: FM029B 005:32.520 005:34.192]

(future tense):

ɡɐɾə [SS: FM029B 005:49.253 005:52.736]

gɛɾə [SS: FM029B 009:59.465 010:02.298]

*Words with less vowel variation.* In the current maternal speech sample, there are also words which do not appear to vary much at all in vowel quality across different tokens of the words. For example, the following words shown in (10) consistently contain [ɔ], whose equivalent in English is the phoneme /ɔ/.

(10)

'topside (beef)'

tofsed [SS: FM029B 002:47.944 002:50.452]

'wrong way' :

ɾəŋwe: [SS: FM029B 012:45.637 012:47.541]

There is little perceptible vowel variation in pronunciations of a number of Gurindji Kriol words, shown in (11), which contain [ɛ], and whose English equivalent also
contains /ɛ/: 

(11)

'then': deŋ [SS: FM029B 006:04.184 006:05.484]


'get': gɛdɛm [SS: FM029B 001:20.422 001:25.437]


Other words which do not vary in vowel from [ɛ] include words which in English have other vowels, such as those shown in (12):

(12)

ɛɲ (Kasey) [SS: FM029B 014:34.521 014:36.332]

kɛmbɛk (come back) [SS: FM029B 009:59.465 010:02.298]

Variation suggesting little or no contrast relative to English. Within other words for which some tokens are transcribed with [ɛ] there is considerable variation in vowel production. In pronunciations of Gurindji Kriol words which are cognate with English words involving /ɜ/ (in Australian English, a non-rhotic dialect), there is variation between [ɛ], which is commonly heard, and [ɜ], which is also sometimes heard. This variation suggests a lack of evidence for a separate /ɜ/ phoneme in Gurindji Kriol. Examples are given in (13):

(13)
‘turn’:


τεʌ [SS: FM029B 006:27.692 006:30.153]

‘turtle’:

τετʌ [SS: FM029B 001:33.936 001:37.279]

Summary of Analysis 1

The vowel variation described in this section is consistent with an analysis of Gurindji Kriol as having five monophthong phonemes /ɪ, ɛ, ə, ɔ, ʊ/, with considerable within-word variation and contextual variation. The analysis presented in this section represents a first description of vowel variation in Gurindji Kriol and provides a basis for exploring how children might learn the sound patterns in Gurindji Kriol as a first language. From a developmental point of view, we now move from the qualitative patterns of sound coocurrence and observations about pronunciation differences for individual words to viewing the input to children in quantitative terms.

Analysis 2: Exploring clusters of vowel phones

We next ask the question: what vowel categories might a child learn, just from exposure to vowel token distributions like those in the current sample of maternal speech? We use cluster analysis to address this question in the next section. We are interested in assessing whether an infant or young child might or might not plausibly learn similar vowel categories on the basis of exposure to the distributions of vowel tokens in Gurindji Kriol maternal speech as we have identified in the linguistic description.
Method

The input to the cluster analysis was the dataset of formant values summarized in Figure 1 and Table 1 above. Prior to running the cluster analysis, the midpoint F1 and F2 values of the vowel tokens in Hertz were transformed into Bark, using the equation from Traunmüller (1997):

\[ \text{Bark} = \left[ \frac{26.81}{1 + \frac{1960}{f}} \right] - 0.53 \]

The rationale for using the transformation was to factor into the cluster analysis the greater sensitivity of the human auditory system to differences between lower rather than higher frequencies in speech.

The cluster analysis procedure that we applied was k-means cluster analysis (using SPSS 18.0). K-means cluster analysis was applied first to two measures for each vowel token (midpoint F1 and midpoint F2), and then to three measures for each vowel token (midpoint F1, midpoint F2, and vowel duration).

Cluster analyses were computed for solutions involving 2 to 9 clusters. (The choice of 9 as the upper number is based on the authors' experience with the phonetic transcription, where 9 vowel symbol categories were reasonably well resolved but not more.) Up to 40 iterations were allowed so that each analysis would converge. Within each solution, the distance from each data point to its cluster centre was calculated, and the sum of squared errors (SSE, the sum of these squared distances) was then calculated for each cluster solution, to be used in the evaluation of the k-means solutions.

The output of cluster analyses provides data which can be interpreted as a guide to the likely number of clusters. Sum of squared error (SSE) can be used as a criterion for choosing between solutions. Solutions involving more clusters generally have a
smaller SSE, as expected, but a flattening in the SSE function can be taken as
evidence for one cluster solution over others. (The rationale and limitations of this
approach to cluster evaluation are discussed by Steinley, 2006).

Results

Figure 2 shows the sum of squared error (SSE) for each cluster solution. The SSE
values for analyses of F1 and F2 data are shown with squares, and for F1, F2 and
vowel duration data with triangles. It is clear from Figure 2 that the addition of vowel
duration to the clustering model has negligible effects on the reduction in SSE for
these data (probably because the speech rate is relatively fast).

In the present data, it is not straightforward to identify a 'flattening' in the functions
so as to choose one cluster solution over others. The line does appear to be flatter to
the right of the data point for the three-cluster solution, or perhaps the four-cluster
solution. There are still sizable reductions in SSE as the number of clusters per
solution increases, however.

-------------

Insert Figure 2 here

-------------

Is it possible to decide between the three- and four-cluster solutions? In these
solutions, all clusters are of comparable size; there are at least no implausibly small
clusters in these solutions, so it is not easy to rule out one of the solutions based on
this criterion. Figures 3 and 4 show the three- and four-cluster solutions, respectively,
plotted in F1 x F2 space. (These cluster solutions are those given to the data from
midpoint F1, F2, and vowel duration.) In each figure, for reference, each vowel token
is plotted with the IPA symbol it was given in the transcription stage.

-------------
We can compare the cluster solutions in terms of how quickly they converged. The three-cluster solution converged at 10 iterations, whereas the four-cluster solution converged at 35 iterations (for F1 and F2 data) or 36 iterations (for F1, F2 and vowel duration data). Based on these data, the most stable cluster solution would seem to be the three-cluster solution.

**Summary of Analysis 2**

The cluster analysis was conducted to address the question of which vowel categories an infant might plausibly learn in Gurindji Kriol, just from exposure to the distribution of vowel tokens in perceptual-acoustic space. Supposing that an infant were to base their initial vowel categorisation solely on the distribution of vowels in F1 x F2 x vowel duration space, the results of the cluster analysis suggest that an infant might acquire three initial vowel categories.

This is interesting from the perspective of understanding how children might learn the vowel system in Gurindji Kriol, a language which linguistic analysis suggests contains 5 short vowel phonemes. Due to the distribution of vowel tokens in F1-F2 maternal vowel space, if children infer phonetic categories from peaks in the distribution, they may start with a different number of vowel categories than the descriptive analysis posits for the phonemes of the language as a whole.

**General Discussion**

The present study set out to describe vowel variation within a sample of maternal speech in the mixed language Gurindji Kriol, and to use cluster analysis to evaluate the vowel categories that an infant exposed to Gurindji Kriol might plausibly learn,
under a version of the distributional learning hypothesis (Anderson, Morgan, & White, 2003; Maye, Werker, & Gerken, 2002). In Analysis 1 we provided acoustic support for the fairly narrow phonetic transcription of the vowels in the sample, quantified the relative frequency of different phones in the transcribed sample (Table 1), and described the patterns of vowel variation within words and between phonetic contexts in the sample. It was clear from Table 1 that there were large frequency disparities between vowel phones in the sample, which were striking in view of the five vowel phonemes suggested by minimal and near-minimal pairs in Gurindji Kriol. In particular, the frequency of [ɪ] and [ɐ] is very high (each is around 25% of all vowel tokens), whereas the frequency of [ɛ] and [ʊ] is lower (each at 11-12% of vowel tokens) and the frequency of tokens of [ɔ], centring on the fifth vowel phoneme in the descriptive analysis is very low, at (2% of all vowel tokens).

In Analysis 2, we presented the results of k-means cluster analysis to try to model what an infant might learn solely from the distributional of vowel tokens in perceptual-acoustic space (involving midpoint F1, F2, and vowel duration). The results of the clustering suggest that a three-cluster solution is probably the most defensible cluster analysis solution, although the evidence for this solution over solutions involving more clusters is not overwhelming. With that in mind, we consider first what the implications are for learning Gurindji Kriol, and languages like it, if we assume that something like acoustic clustering is used by infants to form early categories. Second, we consider what interpretational alternatives are available by rethinking various assumptions.

If we assume that infants learn distributionally by doing something akin to cluster analysis on static formant frequency and vowel duration data, the implications from the results of Analysis 2 seem to be that an infant exposed mainly to Gurindji Kriol
that is similar to the maternal speech sample analysed here would probably learn three initial vowel categories. Judging from the clustering outputs, these three early categories would seem to be centred around [ɪ], [ʊ], and a lower vowel category incorporating the vowel qualities [æ, ə, ɛ, θ]. Although the minimal and near-minimal pair evidence points to five vowel phonemes in Gurindji Kriol, due to the relative infrequency of [ɛ] and in particular [ɔ], phonemes in the mid front and mid back vowel space may take the infant longer to resolve. It is perhaps not a coincidence that the U-shaped development found for Catalan-Spanish bilingual infants has, in vowels, come from mid vowels [ɛ]-[e] and [o]-[u]. We could speculate that infants develop peripheral vowel categories earlier, perhaps especially when native language mid vowels lie in vowel systems with many vertical layers (e.g. Catalan) or within relatively vertically small vowel spaces as has been found for Australian Aboriginal vowel systems, including modern contact varieties, by Butcher (1994).

On this view of development, how long would perceptual reorganization take? How long it might take for an infant to progress from three vowel categories to a larger vowel system implied by the patterns of minimal pair contrast is hard to estimate from the current maternal sample. The current sample (at just over 20 minutes) represents a very small slice of experience in an infant's life. When the size of real input over time is considered, distributional learning may allow an infant to develop categories even for overlapping or tightly crowded mid vowels, and this may be what is observed within the first year of life by Bosch and Sebastián-Gallés (2003, 2009).

The current data do not, however, include any perceptual evidence from adults or children about perceptual categories for vowels in Gurindji Kriol, which demands
caution in assuming that the developmental progression is towards a particular number of adult vowel categories. The current data comprise distributional linguistic analysis (Analysis 1) and maternal production data from one speaker (Analysis 2). In light of this it is important to note that we do not know how many vowels are perceptually grouped by adult speakers of Gurindji Kriol. We also do not have data on the relative sharpness with which adult speakers of Gurindji Kriol might perceive various vowel categories in their language. It is possible, for example, that mid vowels remain less sharply identified by adults due to later emergence of this category in response to weaker acoustic and lexical evidence within Gurindji Kriol.

Were the developmental progression to be from three to five vowel phoneme categories, a reviewer points out that the learner would face a reclassification task in which the learner would have to form each of the two new categories by reassigning members of the adjacent categories. So, roughly, /ɛ/ members would come from some of the old /ɪ/ and /ɐ/, for example, and /ɔ/ from some old /ʊ/ and /ɐ/. This would seem to be a more demanding reclassification challenge than a split of one earlier category into two, as the reviewer points out, such as may occur for /n/ and /ŋ/ for Filipino-learning infants (Narayan, Werker, & Beddor, 2010). If reassignment from two adjacent categories into a new, third, middle vowel category is required, it is important to consider how this could happen. One possibility is that the vowels to be reassigned could be gradually identified, on a word-specific basis, as the infant / child becomes more familiar with the range of acceptable variation for vowels in specific words (cf. the suggestion by Swingley, 2009, about the role of early lexical knowledge). This idea probably predicts that words which the infant hears more frequently would tend to be reassigned earlier.
Perhaps, however, the picture of vowel learning that is suggested by the analyses in this paper is misled by various assumptions. The interpretation of the results of the cluster analysis is constrained by two things: firstly, a number of sources of variability not considered in the current data, which would tend to simplify the categorization problem, and secondly, the particular version of the distributional learning hypothesis explored in this paper. The present results have in common with other attempts to classify vowels (e.g. Hillenbrand et al, 1995) that they are limited by the parameters not included in the analysis. The present data to which cluster analysis was applied comprised midpoint F1, midpoint F2, and vowel duration values for each vowel token. This leaves out many well recognised aspects of systematic vowel variability, including F0, higher formants such as F3, and dynamic spectral information about the vowels. The latter was found useful in vowel classification in the study of American English citation vowels by Hillenbrand et al. (1995), and there is evidence that such information can be used by children (Nittrouer, 2007). Although research into context effects with infants and children (e.g. Nittrouer & Miller, 1997) is limited compared with studies of adults, the disambiguating potential of segmental and prosodic context should probably also not be underestimated for young learners. To sum up, many aspects of vowel variability could potentially be drawn on by the infant, and to the extent that they were used perceptually, would tend to make it easier for them to distinguish vowels in the input than the results of the present analysis suggest. In this way the infant might begin with a set of phonetic categories that are more similar to the adult categories, and the process might be a matter of gradually increasing efficiency and accuracy in using input variability in processing, over infancy and childhood.
The present study has explored the vowel categories that an infant learning Gurindji Kriol might plausibly learn on a (simple) distributional learning account. It is important to note that we have been concerned only with the distributions of vowels in maternal speech input and the way in which these appear to cluster, and have not tested infants' perception. The possibility that in Gurindji Kriol infants initially acquire three vowel categories remains a hypothesis for future research.
References


Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops:
acoustical measurements. *Word*, 20, 384-482.


Table 1. Formant values at vowel midpoint and vowel duration for transcribed phones

<table>
<thead>
<tr>
<th>Phone</th>
<th>F1 (Hz)</th>
<th>F2 (Hz)</th>
<th>Duration (ms)</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>æ</td>
<td>705.61</td>
<td>1994.31</td>
<td>97</td>
<td>33</td>
<td>3.8</td>
</tr>
<tr>
<td>e</td>
<td>699.37</td>
<td>1622.31</td>
<td>94</td>
<td>215</td>
<td>24.5</td>
</tr>
<tr>
<td>ø</td>
<td>487.36</td>
<td>2277.65</td>
<td>111</td>
<td>15</td>
<td>1.7</td>
</tr>
<tr>
<td>ø</td>
<td>493.38</td>
<td>1648.64</td>
<td>58</td>
<td>108</td>
<td>12.3</td>
</tr>
<tr>
<td>ε</td>
<td>562.86</td>
<td>2033.17</td>
<td>91</td>
<td>102</td>
<td>11.6</td>
</tr>
<tr>
<td>ə</td>
<td>491.71</td>
<td>1553.49</td>
<td>85</td>
<td>6</td>
<td>0.7</td>
</tr>
<tr>
<td>i</td>
<td>436.93</td>
<td>2300.38</td>
<td>67</td>
<td>244</td>
<td>27.8</td>
</tr>
<tr>
<td>o</td>
<td>567.26</td>
<td>1103.86</td>
<td>145</td>
<td>8</td>
<td>0.9</td>
</tr>
<tr>
<td>ɔ</td>
<td>587.97</td>
<td>1246.46</td>
<td>93</td>
<td>21</td>
<td>2.4</td>
</tr>
<tr>
<td>u</td>
<td>421.03</td>
<td>1843.79</td>
<td>64</td>
<td>17</td>
<td>1.9</td>
</tr>
<tr>
<td>u</td>
<td>446.15</td>
<td>1216.76</td>
<td>71</td>
<td>110</td>
<td>12.5</td>
</tr>
</tbody>
</table>
Figure captions

Figure 1. Acoustic distribution of transcribed vowel phones in the Gurindji Kriol maternal speech sample

Figure 2. Sum of squared errors (SSE) for k-means solutions involving 2 to 10 clusters

Figure 3. Three-cluster solution

Figure 4. Four-cluster solution