CHAPTER THIRTEEN

ANGLOPHONE PERCEPTIONS OF ARABIC SYLLABLE STRUCTURE

Azra Ali, Michael Ingleby and David Peebles

In this chapter, we use a novel empirical method based on incongruent audiovisual speech data to highlight the mediating role that syllabic structure plays in speech perception. To emphasize the internalization of the syllable, we describe experiments probing the differences between Arabophones and Anglophones in the mental models mediating perception. The two groups respond differently to English and Arabic incongruent speech stimuli, and we argue that the contrasting behavior reflects a basic difference in the internal syllabic structure engaged during perception.

Many studies have shown that a syllable has a hierarchical structure with two main constituents: the consonantal onset and the vocalic rhyme. The theoretical consensus is divided between preferences for an onset-rhyme structure, and for inclinations to use an intermediate mora bearing notions of weight and duration (see, e.g., Ewen and van der Hulst 2001, for a general view; and for uses of moraic structure as a representation of field data on dialect variation Watson 2002 and Kiparsky 2002 and chapter 2, this volume). Most psycholinguistic experiments to test the cognitive reality of the syllable and its constituents in the mental models of humans have taken the onset-rhyme pathway in their analysis of laboratory results. Earliest experiments have involved word games using word-stimuli with concealed parts. The simplest games used monosyllabic words C(C)VC(C) (Treiman 1983, 1985). In most of Treiman's work, participants were asked to coin a new word from given pairs of words. The participants created new words by splitting given words at perceived inner boundaries, usually located between onset and rhyme positions. They took either the onset of the first word with the rhyme of the second, or the onset of the second and the rhyme of the first. These studies concluded that participants make onset-rhyme partitioning more often than by chance—suggesting the empirical reality of an onset-rhyme boundary in the cognitive word model of participants. Priming experiments in

which an auditory prime stimulus precedes a visual presentation of a target word are also indicative of the reality of this constituent boundary. Experiments with visual prime preceding audio target, and either target or prime masked, yield results in the form of decision-times for target perception. The decision times are sensitive to the location of the mask, and change most notably as the mask location is moved to cross a constituent boundary. Such psycholinguistic experiments are designed on the supposition that perception of the priming stimulus and the separate perception of the target stimulus are mediated by the same structural model of lexical items, even when one is auditory and the other is visual—textual or pictorial (Segui and Ferrand 2002). It is possible that this identity of structure maybe false, and therefore empirical probes that engage only one kind of perceptual processing are especially valuable. We have sought for such a probe, targeting the natural processing that humans use to understand speech when watching audiovisual presentations—for example, 'headshots' in a news broadcast or documentary program.

Our selected probe uses the McGurk effect which is a response to dubbed recordings of natural speech in which the audio and visual channels differ phonetically at a key segment, but remain temporally aligned. The classic McGurk-MacDonald experiment (MacDonald and McGurk 1978, McGurk and MacDonald 1976) focused on audiovisual fusion in onset of CV syllables, as represented by the notation (1) below. The rule represents the outcome of presenting an audiovisual incongruent speech stimulus with labial place of articulation [ba] in the Audio channel (A) and velar gesture [ga] in the Visual channel (V), to a group of participants who reported a fusion (F) percept [da] in 64% of the cases, 9% reported audio syllable [ba] and 27% reported the visual channel [ga] of the stimulus. With different voiceless plosive stimulus as in (2), 50% of the participants reported fusion percept [ta] and 50% of the participants reported the audio syllable [pa] (MacDonald and McGurk 1978: 255).

(1) $_{A}(\text{ba} \parallel \text{ga})_{V} \rightarrow (\text{da})_{F\ 0.64}, (\text{ba})_{A\ 0.09}, (\text{ga})_{V\ 0.27}$

(2) $_{A}(\text{pa} \parallel \text{ka})_{V} \rightarrow (\text{ta})_{F\ 0.50}, (\text{pa})_{A\ .50}$

There is a long history of using incongruent speech in experiments. Something similar to the McGurk effect also occurs in purely audio contexts when there is incongruence between signals presented to left and right ears of participants. Cutting (1975) noted phonological fusion when using dichotically incongruent audio stimuli, in which

the left-ear (L) and right-ear (R) signal differed at a single segment. Dichotic fusion patterns are shared by adults and children in all stages of linguistic development, and can also be represented symbolically as in the following examples (3) to (4).

(3) $_{L}$(leɪ || peɪ)$_{R}$ → (pleɪ)$_{AF}$, (lpeɪ)$_{AF}$, (leɪ)$_{L}$, (peɪ)$_{R}$

(4) $_{L}$(tæs || tæk)$_{R}$ → (tæsk)$_{AF}$, (tæks)$_{AF}$, (tæs)$_{L}$, (tæk)$_{R}$

In (3), the phonologically licit percept 'play' respects the phonotactic constraints of English branching onsets, and was reported even when the temporal alignment favored the phonologically illicit 'lpay'. In (4), however, there are two fusion responses reported by English participants, (tæsk)$_{AF}$ and (tæks)$_{AF}$ and these both have branching codas that are equally licit in English phonotactics.

More recent laboratory work on dichotic incongruence has focused more on so-called migration than on fusion. The classic experiments (Kolinsky and Morais 1993, Mattys and Melhorn 2005) also relate to the notion of an internalized syllabic structure mediating speech perception. The stimulus tends to evoke an illusory percept M, made up of contiguous groups of segments from R and L channels that migrate unchanged into the percept as in (5) (Mattys and Melhorn).

(5) $_{L}$(dɒlmal || kɛ:rfɪn)$_{R}$ → (dɒlfɪn)$_{M}$, (kɛ:rmal)$_{M}$, (dɒlmal)$_{L}$

Rule (5) represents participants reporting migration percept [dɒlfɪn] and [kɛ:rmal], and some participants opting for the left-channel signal [dɒlmal] and the right-channel signal [kɛ:rfɪn]. With dichotically-presented polysyllables, the recombinative illusions with the highest rates occur as migrations of units that are close correlatives of the syllables of traditional linguistic analysis (rather than segments or other units). Such experiments provide confirmatory evidence for the role of syllable in lexical access.

Our use of the illusions elicited in McGurk fusion is analogous to such migration studies, and probes several organizational features of the mental lexicon: empirical correlates of syllabic onsets and codas, the cohesion of morphological affixes, etc. In section 2, we outline evidence that the different fusion rates (frequency of fusion responses) represented in these rules are typical of a tendency for higher rates when the incongruent segment is in the syllabic coda, lower rates at onsets. Furthermore, they show empirically that syllable structure

is indeed part of speakers' mental phonological representation. In sections 3 and 4, we illustrate this by showing how fusion rate trends differ systematically between participant groups of Arabic and English mother tongue when they are given the same stimuli.

Although early research on McGurk fusion with CV syllables was psychophysical in orientation, observation of group response to incongruent data in natural language context provides a direct route to the lexical access mechanisms in human cognition. It is relatively free of metalinguistic factors that might influence word-games with masked and primed stimuli. In addition, it can be given a high degree of ecological validity by placing incongruent stimuli randomly amongst congruent audiovisual stimuli, thus ensuring that participants process incongruent speech using the same mental apparatus that is engaged by congruent speech. In this way one can probe rather directly the internal mediation of speech cognition by structural organization of the mental lexicon. Such directness is important when investigating constituent structure as perceived by different groups.

1 *McGurk Fusion and the Mental Lexicon*

McGurk fusion studies have been accumulating in many laboratories for more than thirty years. The illusion is persistent even when participants are told that the soundtrack does not match the visual lip movements of the speaker. It survives size reduction of the video image, and it is also robust against acoustic noise (Tiippana et al. 2000, and others) and against visual noise (Fixmer and Hawkins 1998). In fact, precise temporal alignment of the audio and visual channel is not needed for experiencing McGurk fusion (van Wassenhove et al. 2007). Fusion responses still occurred even with temporal asynchronies from −30 ms to +170 ms.

McGurk fusion phenomena are also known to survive embedding in many natural languages, occurring amongst speakers of different mother tongues: French (Colin et al. 1998), Dutch (de Gelder et al. 1995), Finnish (Sams et al. 1998), and Chinese (de Gelder et al. 1995). More recent studies have used the McGurk effect to probe the lexical knowledge and semantic processing in Finnish (Sams et al. 1998), in German (Windmann 2004) and in English language (Brancazio 2004; Barutchu et al. 2008). Sams et al. focused on McGurk fusion in real

or nonce words which were presented either in isolation or in a three word phrase context. Windmann used a semantic priming approach, where the isolated incongruent audiovisual stimuli were either semantically coherent or incoherent with a textually-presented prime. Her results showed greater fusion rates for semantically related primes than for semantically unrelated primes. Both Sams and Windmann studies show that lexical effects influence but do not over-ride McGurk fusion.

If fusion is truly a predominantly phonological process, then the fusion found in responses to incongruent consonant segment should also occur in incongruent vowel segments too. This is indeed the case, as shown in studies with Swedish and English vowels (Öhrström and Traunmüller 2004; Ali and Ingleby 2002). In these studies, an incongruent vowel segment was embedded in real words and nonsense CVC syllables which evoked fusion at similar rates to those found in consonant fusion.

2 *Probing Syllable Structure*

With appropriate statistical analysis, experiments on the McGurk effect have been successful in probing syllabic structure. The probe was a set of word stimuli in which an audiovisually incongruent segment was in either an onset site, or a vowel nucleus, or a coda site. The key experimental finding was that the site of the incongruent segment significantly affects the phonological fusion rate. This fact allows one to infer, from fusion rate comparisons, whether a given segment is part of, say, an onset, or of some other syllabic constituent. In this section, we briefly detail some of our earlier work using the McGurk effect to probe syllabic structure in English words, then in sections 3 and 4 we detail a feasibility study working with incongruent segments embedded in Arabic words.

In our earliest fusion studies (Ali and Ingleby 2004), stimuli were restricted to simple monosyllabic words of CVC type. An incongruent consonant segment was embedded either in the onset or in the coda site. In order to generate a stimuli set, word-triples are required; for example, a word in the audio channel, a word in the visual channel, and the expected fusion, that differs only at a single place segment, as illustrated in typical segmental incongruence and qualitative fusion results in (6) and (7).

(6) $_{A}$(**b**eɪt || **g**eɪt)$_{V}$ → (**d**eɪt)$_{F,}$ (**b**eɪt)$_{A}$,… {onset}

(7) $_{A}$(flæ**p** || flæ**k**)$_{V}$ → (flæ**t**)$_{F,}$ (flæ**p**)$_{A}$,… {coda}

Our simple, monosyllabic, word-triple stimuli are listed in Appendix 1. The experimental method and procedure was the essentially same as that used in the bilingual study of sections 3.2 to 3.4, except the words were English and all participants were Anglophones. From the small sample of stimuli tested, quantitatively, fusion rates were greater for consonantal codas than onset consonants (60% and 48% respectively, statistically significant). Our observed fusion rate contrasts were obtained from sequences including congruent audiovisual stimuli—for assurance of ecological validity and of the effects of signal-quality perception inaccuracy. Although we did not control for equal number of voiced and voiceless word-triples, for future work, we have managed to generate a much larger word-triple list with equal number of voiced and voiceless plosives.

In later monolingual experiments, we explored the onset-coda contrast in branching constituents, placing the incongruent segment either in the first or second branch of an onset, or in the coda position, as illustrated by cases (8) to (11).

(8) $_{A}$(**b**reɪz || **g**reɪz)$_{V}$ → (**d**reɪz)$_{F,}$ (**b**reɪz)$_{A}$,… {onset: Cr}

(9) $_{A}$(s**p**eəz || s**k**eəz)$_{V}$ → (s**t**eəz)$_{F,}$ (s**k**eəz)$_{A}$,… {onset: sC}

(10) $_{A}$(kɒ**b**z || kɒ**g**z)$_{V}$ → (kɒ**d**z)$_{F,}$ (kɒ**b**z)$_{A}$,… {coda: Cs}

(11) $_{A}$(kɔːr**p**[1] || kɔːr**k**[1])$_{V}$ → (kɔːr**t**)$_{F,}$ (kɔːr**p**)$_{A}$,… {coda: cC}

Our stimuli with branching constituents are also listed in Appendix 1. Again, the experimental method was the same as in the previous CVC experiment. Some researchers might argue that the stimuli in case (9) are questionable, that they are not branching onsets because word initial [s]+stop clusters violate the Sonority Sequencing Principle (SSP; Clements, 1990). The SSP states that the segments of a syllable are arranged in a way that their sonority increases, from the beginning of

[1] Articulated rhotically by a Scottish speaker, to avoid problems with /r/ amongst English speakers.

the syllable onset to the nuclear peak, and decreases afterwards to the end of the syllable. Therefore, SSP governs the permissible sequences of consonants within syllables. The permissible ordering follows the so called Giegerich scale (Giegerich 1995). Kaye (1992), and Pan and Snyder (2004) have taken the view that SSP-violating word-initial sC clusters are structurally different from other branching onsets. Kaye proposes that word initial sC is binary, with [s] extrasyllabic, and has attempted to support this with evidence/examples from Italian, Portuguese, Ancient Greek, and English. Our motivation for keeping the sC as nominal branching constituents was to see whether or not the fusion rates are the same as onset Cr as in case (8). There are some similar concerns about [s] and [z] in the second branch of a branching coda. The concern relates to the polymorphic nature of English plurals, with or without voicing of [s] to form [z] via voicing harmony. Perhaps these segments might be considered extra-syllabic or otherwise exceptional on SSP grounds.

Our measurements, however, showed that incongruity sited in the first branch of an onset (Cr) influenced fusion much as in the second branch (sC): C-fusion rates for onset Cr and sC were 27.0% and 27.3% respectively, not significantly different. Also, as in the onset case, the effect of the plural markers in the first branch of coda constituent (Cs) was much the same as for second branch constituent (cC): C-fusion rates were 40.1% and 40.9% respectively. The results confirm a significant main effect in fusion rates between branches of coda and branches of onsets, but no significant effect within branches of onsets or within branches of codas.

2.1 *Perceptual Place Cues in Onset and Coda*

There might be some concerns about plosive fusion perception in the coda position, because plosives, (especially phonemes [p] and [t]), are generally hard to distinguish on the basis of their burst spectrum. Usually, formant transitions in syllable onsets are more reliable cues to consonant place of articulation (POA) than coda transitions (Liberman et al. 1967; Wright 2004). Wright (2001) tested intelligibility rate of consonants in onsets and in codas in presence of variable acoustic signal-to-noise. Misperceptions of consonant POA were made by participants, but mainly in noisy conditions, and POA misperceptions were more frequent for consonants in the coda position than in the onset position. Thus, it is important to check whether the alveolar fusions perceived in coda by many of our participants were genuine

and were not a confusion of place contrasts (between [p] and [t] and between [b] and [d]), more frequent in coda than onset consonants.

We argue that the alveolar fusion reported by the participants in our studies were genuine perceptions rather than matters of confusion. Our argument is based on four points. Firstly, our experiments were not presented in presence of noise at the levels used by Wright to elicit confusion (our audio signals were presented binaurally over headphones rather than through speakers). Secondly, with congruent audiovisual stimuli which were embedded as controls amongst incongruent stimuli, our participants achieved 100% accuracy for place recognition. Thirdly, our participants were operating audiovisually with visible lip gestures supporting place perception. And fourthly, Clement and Carney (1999) showed that with incongruent audiovisual speech stimuli, the visual signal is favoured more when the audio signal is degraded. This is consonant with the findings of Sumby and Pollack (1954)—that the accuracy rate in presence of noise is greater for audiovisual speech stimuli than for audio only modality—and the findings of Inverson, Bernstein and Auer (1998)—that in visual mode only and in audiovisual modality visual, POA cues are clearly distinguishable. The subject of visual POA cues has been related to external lip-shapes corresponding to labial, alveolar and velar mouth gestures. Shapes have been sketched (Harris and Lindsey 1995) and proffered as the visual cues of phonological elements of speech. They classify these visual cues as a wide mouth opening for element A, a rectangular, tight lip-shape for element I and a rounding of lips for element U (Harris and Lindsey 1995). Via such cues, the different elements can be detected both in combination and in isolation within phonetic segments. A phonological framework, whose subsegmental primitives are, to an extent, both audible and visible, is ideal for modeling audiovisual speech phenomena such as McGurk fusion as illustrated in Ingleby and Ali (2003).

3 *Codaless Languages and Arabaphone Perception of Arabic Syllable Structure*

In a language without codas, the quantitative distinction between fusion rates in codas and onsets obviously cannot persist. The case of Arabic is especially interesting. The Arabic tradition of Sybawaih, on which the (phonetic) Arabic alphabet is founded, uses CV units only, symbolised orthographically by a consonant bearing a vowel diacritic.

This is also visible in the Quranic script, where consonants are clearly marked with vowel diacritics and where there are no vowels, a *sukuun*, a circle diacritic denoting silence, is marked above the consonants. The Western tradition of classical scholars, treating Arabic like Latin and Greek, postulated that there are CVC, CVVC, CVCC syllables, and assigned to consonant segments a notional 'coda' or 'onset' label. Yet others have maintained that Arabic words are built on one type of syllable: CV (Guerssel and Lowenstamm, 1996). The CV view is developed at a level of phonological abstraction above what is customary when dealing with the phonetics/phonology interface, but newer findings closer to the interface are emerging.

The most interesting findings were made by Baothman and Ingleby (2002), confirming empirically in a large corpus of spoken Saudi Arabic, that, unsurprisingly, it has no branching onsets but more significantly that sukuun has a complex articulatory presence. They found evidence in recordings and spectrograms that, far from being merely an orthographic device, a phonologically active correlative of *sukuun* is present within consonantal clusters such as [bʒ]. Its presence is revealed in the durations of clusters, which are longer by a short-vowel interval than the sum of the consonant durations when they are not clustered. In addition, the 'phonological sukuun' triggers the phonological processes of consonant devoicing and vowel epenthesis. For example, in the final cluster of Arabic بَحْر ([baħr] = 'sea') a schwa vowel [ə] can be heard and also detected as vowel material in spectrograms. This articulatory evidence indicates that *sukuun* is a cognitive reality for Modern Standard Arabic (MSA) speakers. In section 5, we discuss this further.

We suggest that if the mental lexicon of Arabophones has no codas, as in Sybawaihan tradition and in the work of Baothman and Ingleby, then there should be no statistical difference in fusion rates measured for onsets and 'notional codas' (consonants labeled as codas in the Western tradition). The issue can be tested via the perceptions of native speakers using Arabic word stimuli with an incongruent phonetic segment that would be a notional 'coda' in the Western tradition, but an onset in the Sybawaih view. Consonants in word-medial (e.g., C_2, Figure 1) position of English words can be either codas (as in 'habit', 'hacker', 'haddock', 'halo') or onsets (as in 'habituate', 'hallo', 'harangue'). The onset cases are less frequent and only occur when the second syllabic nucleus carries the primary stress. Thus, Anglophones will tend by habituation to perceive word-medial consonants in nonce

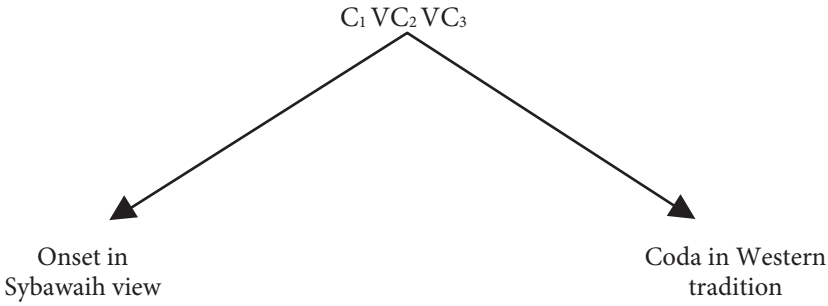

Figure 1 Possible syllabification

words or Arabic words as codas. The Arabic examples that we have used have in some cases second nuclei that are long vowels (/habaa/ etc.). An Anglophone reading these will mispronounce and place primary stress on the second nucleus, making the preceding consonant onset-like. Our participants, however, were not given this opportunity because they received the words audiovisually from an Arabic speaker placing stress on the first short nucleus for morphological reasons.

Thus, in our feasibility study, we have two predictions:

1. In Sybawaih view, we predict that Arabophones should show similar fusion rates in all 3 consonantal positions, thus, indicating a no onset-coda distinction.
2. In Western tradition, we predict that Anglophones should show differences in fusion rates: greater fusion in C_2 and C_3 than in C_1, so C_2 is likely to behave as notional coda than as an onset.

3.1 *Phonological Context of Arabic Consonants*

Arabic has many distinct dialects, but there is a *koiné dialektikos* used by many educated native speakers in business, media and education. It is close to the classical Arabic whose elocutionary form finds expression in the *tajweed* of Quranic scholars, and is usually referred to as Modern Standard Arabic (MSA). Our speaker data is taken from users of this MSA. The speakers do not share mother-dialect with all of the participants in our perception experiments, but speakers and participants share a set of phonological segment contrasts even though they do not realize them phonetically in the same way. For example, users of MSA articulate the orthographic 'qaaf' as voiceless uvular [q] while

in Yemen, for example, a voiced velar [g] is common. The place of articulation contrasts used in our English stimuli are lacking in MSA, which has:

- no voiceless labial [p] except word-finally as output of a devoicing process, ruling out [p]-[k] incongruent pairs;
- no voiced velar [g], ruling out [b]-[g] incongruence pairs;
- two velar places, a front velar [k] and a back velar or purely uvular [q] from the letter 'qaaf';
- pharyngeal (or emphatic) place contrasts that complicate the pairing of incongruous segments—whereas [b] has no pharyngeal variant, [d] and [t] have both alveolar forms د and ت, and pharyngeal forms ض and ط, representable in IPA as [dˁ] and [tˁ] respectively.

In addition to these place and manner complications, there are contextual issues different from those of English. All Arabic consonants can be geminated, and this complicates the simple classification of consonants as onset or coda. One needs to experiment with geminates as well as singleton consonantal, but that is for later studies.

3.2 *Participants and Experimental Design*

Ten participants took part in the experiment, with an age range between 24 to 31 years, all university graduates familiar with MSA. Their mother tongue was Arabic in the dialects of Jordan, Kuwait, Oman or Israel/Palestine. The participants did not have any specialized knowledge of linguistics or psychology and were from computing and business disciplines.

A small lexicon of Arabic words with appropriate place contrasts are shown in examples (12) to (14). A full list of Arabic word-triples is detailed in Appendix 1. To avoid problems with the [k]–[q] contrast, we favoured words with [q] whenever possible. In the interest of assuring ecological validity, a sample of congruent stimuli was also included in the experiment, and all words were real meaningful words. The stimuli were constructed from the speech of a male speaker (from Morocco) and a female speaker (from Syria), both native Arabic speakers accustomed to MSA, with postgraduate education and age in the range 20–40 years.

(12) $_{A}$(balla || qalla)$_{V}$ → (dalla)$_{F}$,…. {C-initial}

$_{A}$(بلّ || قلّ)$_{V}$ → (دلّ)$_{F}$,….

(13) $_{A}$(ħabaa || ħaqaa)$_{V}$ → (ħadaa)$_{F}$,.... {C-medial}
$_{A}$(حَبَى || حَقَىَ)$_{V}$ → (حَدَا)$_{F}$,....

(14) $_{A}$(nahab || nahaq)$_{V}$ → (nahad)$_{F}$,.... {C-final}
$_{A}$(نَهَبْ || نَهَقْ)$_{V}$ → (نَهَدْ)$_{F}$,....

3.3 *Creating incongruent stimuli*

Video recordings were done inside a quiet laboratory using a standard 8 mm digital Sony Camcorder with a built-in microphone for audio. The speakers uttered each word twice, to enable best selection from the two in terms of audibility and visibility. The speakers produced the words with clear voice to prevent effects of coarticulation of the sort that might affect place perception. The video recording was spliced into individual word recordings. These splices were stored in a standard *.avi file format with a frame size 640 × 512 pixel and 25 frames per second.

For the creation of incongruent stimuli, standard editing software (Adobe Premier 5.5) was used. Recordings differing in one phonetic segment were paired: for example [baal] (video 1) with [qaal] (video 2). The audio channel from video 1 was imported into video 2 and aligned with the audio and visual channel of video 2. The experimenter made fine judgments of proper alignment manually, after previewing the video clip. After alignment, the audio of [qaal] (video 2) channel was erased. The resultant video was thus [baal] in the audio channel aligned with visual lip movements saying [qaal]. All the stimuli (congruent and incongruent audiovisual stimuli) were stored in a standard *.avi file format using the 640 × 512 pixel frame size at 25 frames per second.

3.4 *Instruction Sheet, Response Form, and Procedure*

The participants were provided with report forms on which to record 'what they thought the speaker was saying' when receiving an experimental stimulus. The forms listed options corresponding to the words in audio and video channels, the expected fusion response, two random words and finally, a space to write in any word not explicit in the list. The experiment was carried out in a controlled laboratory with minimal background noise. Participants sat about half a meter from the 17" monitor screen and used headphones connected to the computer to listen to the audio. Using headphones maximized the audio

signal rather than presenting the audio signal via speakers which can be degraded when it reaches the participant's ear. Participants were simply asked to wear the headphones provided, watch the video and decide what the speakers were saying. There was no time limit set and participants were not given any feedback.

3.5 *Results*

Firstly, the data for congruent stimuli were analyzed, revealing that 94% of the participants accurately perceived what the speaker was saying, eliminating poor vision or hearing as a factor influencing results. Secondly, group averages for fusion responses to incongruent stimuli were compiled, which revealed that fusion rates were similar in all three consonantal positions. Since only a small set of Arabic stimuli were tested, we used exploratory categorical statistical analysis only, which showed no main effect of consonantal site ($\chi^2 = 4.083$, $df = 2$, $p = .130$). Since C-medial and C-final consonants are 'notional codas,' fusion rates for onsets vs. notional codas were, 61% and 69% respectively. Statistically again showing no significant main effect of consonantal site ($\chi^2 = .722$, $df = 1$, $p = .395$).

Given that lexical access to Arabic words by Arabophones is performed without onset-coda distinctions, the question of why such distinctions are made by Western scholars arises. Perhaps the teaching of 'classical Arabic' by Western scholars, who use notions of syllable with onset and rhyme from their own and classical languages, is a cultural imposition. To explore the suggestion further, with the idea that it may originate in the cognition of Westerners, we put Arabic stimuli as nonce words to a group of Anglophone participants with no knowledge of Arabic.

4 *Anglophone Perception of Arabic Syllable Structure*

The aim of giving Arabic stimuli to Anglophones was to investigate whether or not, from internalized lexical access models learned from their English mother tongue, they imposed onset-coda distinctions on segments embedded in unfamiliar Arabic speech.

4.1 *Stimuli, Response Form and Procedure*

We put the same Arabic stimuli from our Arabophone experiment to Anglophones. We gave participants an open choice when reporting

their percepts. The response forms consisted of words which were transliterated from Arabic into an English equivalent, for example, أَرَابْ [ʔaraab] was transliterated into English as <aaraab> so that our participants unfamiliar with IPA could respond. Our transcriptions ignored the distinction between the glottal stop, hamza (ء) and the glottal fricative, ayn (ح), rendering both by the textual apostrophe used for English dialect words like ['appen] and [da'abase]. The forms included options corresponding to the words in audio and video channels, the expected fusion response, two random words and a space to write in any word not explicit in the list. The rest of the procedure was exactly the same as Arabophone experiment (Section 3.4).

Eleven participants, 4 females and 7 males, took part in the experiment. All were native British English speakers, with an age range between 21 to 47 years. They had no specialized linguistic or psychology training, nor any knowledge of the Arabic language, nor any experience of visiting any Arab country.

4.1 *Results*

Firstly, the data for congruent stimuli were analyzed, revealing participants with no knowledge of Arabic are capable of reporting correctly to the congruent stimuli. In fact 93% (compared to 94% for Arabophones) of Anglophone participants accurately perceived what the speaker was saying, eliminating poor vision or hearing as a factor influencing results. Secondly, fusion rates were greater in C-medial and C-final than in C-initial position. Statistically, there was a significant main effect for consonantal site ($\chi^2 = 19.898$, $df = 2$, $p < .000$). Fusion rates for onsets vs. notional codas were, 38% and 72% respectively. Statistically, again showing a main effect for consonantal site ($\chi^2 = 15.063$, $df = 1$, $p < .000$). In summary, glancing at Table 1, the fusion rates for Anglophone participants for English and Arabic stimuli were very similar for onsets (40% and 39% respectively), but only slightly different for codas (65% and 73% respectively). Our feasibility study appears to show that Anglophones treat Arabic syllable structure similar to their native language, thus showing an onset-coda distinction.

In later experiments, we put the same Arabic stimuli to Anglophones who had been learning Arabic for at least 3 years on a part-

time basis. The results are summarized in Table 1, and we discuss this in section 5.

Table 1. Average fusion rates at coda and onset sites

Stimuli and Participants	Onsets	Codas
1. English stimuli*—Anglophone participants	40%	65%
2. Arabic stimuli—Arabophone participants	61%	69%
3. Arabic stimuli—Anglophone participants	39%	73%
4. Arabic stimuli—Anglophones learning Arabic	63%	68%

* average fusion rates based on monosyllabic and polysyllabic words.

Of course, within the limitations of small-scale experimentation these results may or may not be replicable in a wider sample. Power analysis using G*Power (Faul et al. 2007) suggests that they are replicable: the effect of size at 0.19 turns out to be not a major factor in the results.

5 *Discussion*

The primary purpose of this chapter was to show that McGurk fusion can be used as a tool for probing internalized syllable percepts in the mental lexicon and determining their structure in two exemplar languages. This was achieved by measuring the variation of fusion rate with the context in which an incongruent consonant speech segment is embedded. Our studies from English and Arabic define a basis for empirical measurement of the syllable context; in particular it allows codas and onsets of syllables to be distinguished empirically.

There are different vulnerabilities to fusion in English; codas are more prone to fusion than onsets and are statistically significant and are robust in monosyllabic words (non-branching and branching constituents). This robustness means that fusion patterns can be used to test hypotheses about whether or not a language has the same onset-nucleus-coda patterns as English, a phonologically interesting question, given that theorists sometimes claim that certain languages are made up entirely of codaless syllables.

Although no-coda languages are known, Arabic syllabic structure is sometimes contested. The nature of the contest is summarized in Section 3: an historical Eastern CV tradition at the heart of standard orthography, versus an external Western scholarly tradition inspired by the syllabification of Latin and Greek. In more recent times, Baothman and Ingleby (2002) have developed a CV model of Arabic speech patterns using element phonology and a codaless constituent structure. It represents all the known coarticulation processes of Arabic and leads to a stress-prediction algorithm that is much simpler than those based on the syllabification of the Western classical tradition. More recent studies have shown vowel epentheses between consonant clusters in Lebanese and Palestinian Arabic (Gouskova and Hall 2007) and in Moroccan Arabic (Ali et al. 2008) that fit a CV model of phonological process. This tilts the contest to favor the Eastern tradition.

In this chapter, we used incongruent stimuli to open up the contest to experimental test. The test was to seek the segments that show too much fusion at all consonantal sites to be onsets: if the search fails, then there are no codas in the sample tested, which should be chosen to be representative of the language under test. In the first experiment (Arabic stimuli put to Arabophones), incongruent consonant segment was at either an onset or a 'coda' position (word-medial or word-final). When we put these stimuli to our Arabophones, results showed that there were no significant differences in fusion rates between onsets and 'codas'. This adds to the growing evidence that Arabic is a CV language, where syllables are coda-less not only at the underlying level, but also at the surface level, and also at the perception level. Of course the no onset-coda difference of fusion rates is also compatible with a VC phonology, but the absence of word-initial vowels in classical Arabic and several (but not all) modern dialects pushes the balance strongly towards CV. The counterbalancing cases have been reviewed by Kiparsky (2003) from the point of view of moraic theory, but they could also be thought of as the output of phonological processes favored by a dialect group, the processes operating on an underlying onset-rhyme CV structure. In the second experiment, we put the same Arabic stimuli to Anglophones with no knowledge of Arabic. These are nonce words to Anglophones and they showed an onset-coda distinction. They showed a similar perceptual pattern to our English syllable studies—higher fusion rates for 'codas' than for onsets. This indicates that in the mental models of Anglophones codas do exist and they are clearly distinguished from onsets. Thus, we have shown empirically that syllable structure is indeed part of speakers'

mental phonological representation: Anglophones show an onset-coda distinction, whilst Arabophones do not.

It is possible that the results can be interpreted differently, considering that lexical access and speech processing in Arabic may be root based. The root consonants are not adjacent to one another in all words that share a root, e.g., root *k-t-b* 'write' is shared by the verb form '*kataba*' ='he wrote' and derivatives like '*yaktab*'= 'he is writing'. In other words, roots are abstract and discontinuous morphemes, stored in the lexicon with no syllable structure. They are given a syllable structure when they are associated with prosodic templates to form words. But, in English, morphemes are stored with their syllable structure and this can be seen as the difference between Semitic languages and Indo-European languages. To some extent this was demonstrated, for example, by Idrissi et al. (2008) with an aphasic bilingual Arabic-French speaking adult that Arabic consonantal roots are abstract morphemic units rather than surface phonetic units. However, we believe that once at the phonological component, the word pattern follows a CV structure, with no distinction between onset and coda. The experimental work presented in this chapter supports this.

In our most recent experiment, we put the same Arabic stimuli to Anglophones who had been learning Arabic for at least 3 years on a part-time basis. Surprisingly, the structure of Arabic in their mental representation shifted towards a CV pattern; showing a no onset-coda distinction (see Table 1, point 4).

In this chapter, Arabic words were used; these were real words for speakers of Arabic, but nonce words for the English speakers with no knowledge of Arabic. In contrast, for future study, we aim to use nonce Arabic words that would be put to Arabophones; our hypothesis is that Arabic speakers will still show a similar pattern to their own language, a no onset-coda distinction.

Finally, we would like to emphasize that experimental and phonetic analysis on Arabic roots and syllable structure has been very limited to date. Only few experimental studies have probed the mental lexicon (Idrissi et al. 2008, Prunet et al. 2000, Boudelaa and Marslen-Wilson 2004, and our own studies presented in this chapter). Few phonetic studies have begun to show phonologically active *sukuun* and vowel epentheses in consonant clusters (Baothman and Ingleby 2002, Guskova and Hall 2007, Ali et al. 2008). Recently, Rosenhouse (2007, 131) stated, "what can be hoped for the future is to further develop phonetic/phonological studies in order to learn more about the system

of the Arabic language". We would like to add that experimental work is also essential because it can probe rather directly the internal mediation of speech cognition by structural organization of the mental lexicon.

6 *Conclusion*

In this chapter, we have shown from the statistics of response to incongruent speech stimuli that internalized syllables do exist. We have further shown that internalized English syllables have a structure in which the coda and onsets are distinct. An abiding pattern in all the context-embeddings that we have investigated in this chapter is that fusion rate patterns and the structural features inferred from them remain significantly different for English and Arabic. In English, codas and onsets embedded in monosyllabic words, are two distinct entities whilst there is only a single consonantal entity in Arabic.

In the extended study, we continue to use the McGurk effect embedded in Arabic words, with an aim of generating a large corpus of incongruent audiovisual stimuli. But we are also motivated by using the syllable migration paradigm for probing Arabic consonantal roots as well as the syllabic structure. We have also begun to investigate contrasting fusion rate patterns for singleton and geminate Arabic consonants. The latter are phonemic in Arabic and very common, whereas in English, though textual gemination is a common orthographic feature, truly phonemic gemination is rarer. When attested, it is a product of collisions—morphological (e.g. 'unknown' and 'soulless'), or cross-word for example 'big game' or 'top post' or across phrase boundaries, for example, 'Jack, cutting in, said…' or 'Pop, posing a question, stood…' etc.

References

Ali, Azra N. and Michael Ingleby. 2002. Perception difficulties and errors in multimodal speech: The case of vowels. In *Proceedings of the 9th Australian International Conference on Speech Science & Technology*, edited by C. Bow, 438–443. Canberra: Australian Speech Science and Technology Association (ASSTA).

——. 2004. Probing cognitive mental models of speech using the McGurk effect. In *The proceedings of the 4th International Conference on the Mental Lexicon*, 44.

Ali, Azra N., Mohamed Lahourchi and Michael Ingleby. 2008. Vowel epenthesis, acoustics and phonology patterns in Moroccan Arabic. In *Interspeech-2008*, 1178–1181. Causal Production: Australia.

Baothman, F. and Michael Ingleby. 2002. Representing coarticulation processes in Arabic speech. In *Perspectives on Arabic Linguistics XVI*, edited by Sami Boudelaa, 95–102. Current Issues in Linguistic Theory 266, Amsterdam: John Benjamin Publishing Co.

Barutchu, Ayla, Sheila G. Crewther, Patricia Kiely, Melanie Murphy and David P. Crewther. 2008. When /b/ill with /g/ill becomes /d/ill: Evidence for a lexical effect in audiovisual speech perception. *European Journal of Cognitive Psychology* 20: 1–11.

Boudelaa, Sami and William Marslen-Wilson. 2004. Abstract Morphemes and lexical representation: The CV-Skeleton in Arabic. *Cognition* 92: 271–303.

Brancazio, Lawrence. 2004. Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 30: 445–463.

Clement, Bart and Arlene Carney. 1999. Audibility and visual biasing in speech perception. *Journal of Acoustical Society of America* 106: 2271.

Clements, George. N. 1990. The role of the sonority cycle in core syllabification. In *Papers in laboratory phonology 1: Between the grammar and physics of speech*, edited by John Kingston and Mary Beckman, 283–333. NY: Cambridge University Press.

Colin, Cecile, Monique Radeau and Paul Deltenre. 1998. Intermodal interactions in speech: a French study. In *Auditory-Visual Speech Processing (AVSP'98)*, edited by Denis Burnham, Jordi Robert-Ribes, and Eric Vatikiotis-Bateson, 55–60. Terrigal-Sydney, Australia. http://www.isca-speech.org/archive/avsp98/

Cutting, James E. 1975. Aspects of phonological fusion. *Journal of Experimental Psychology: Human Perception and Performance* 1: 105–120.

Ewen, Colin J. and Harry van der Hulst. 2001. *The phonological structure of words*, Cambridge: Cambridge University Press.

Faul, Franz, Edgar Erdfelder, Albert-Georg Lang and Axel Buchner. 2007. G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39: 175–191.

Fixmer, Eric and Sarah Hawkins. 1998. The influence of quality of information on the McGurk effect. InAVSP'98: *Proceedings of the International Conference on Auditory-Visual Speech Processing*, edited by Denis Burnham, Jordi Robert-Ribes and Eric Vatikiotis-Bateson, 27–32. Terrigal, Australia.

Gelder, Beatrice de, Paul Bertelson, Jean Vroomen and Hsuan Chin Chen. 1995. Interlanguage differences in the McGurk effect for Dutch and Cantonese listeners. In *Eurospeech-1995*, 1699–1702. http://www.isca-speech.org/archive/eurospeech_1995/

Giegerich, Heinz. 1992. *English phonology*. Cambridge: Cambridge University Press.

Gouskova, Maria and Nancy Hall. 2007. Levantine Arabic Epenthesis: Phonetics, Phonology and Learning. Presented at Variation, Gradience and Frequency in Phonology Workshop, Stanford University.

Guerssel, Mohand and Jean Lowenstamm. 1996. Ablaut in Classical Arabic measure I active verbal forms. In *Studies in Afroasiatic grammar*, edited by Jacqueline Lecarme, Jean Lowenstamm and Ur Shlonsky, 123–134. The Hague: Holland Academic Graphics.

Harris, John and Geoff Lindsey. 1995. The elements of phonological representation. In *Frontiers of phonology: atoms, structures, derivations*, edited by Jacques Durand and Francis Katamba, 34–79. Longman: Harlow, Essex.

Idrissi, Ali, Jean-François Prunet and Renée Béland. 2008. On the Mental Representation of Arabic Roots. *Linguistic Inquiry* 39: 221–25.

Ingleby, Michael and Azra N. Ali. 2003. Phonological Primes and McGurk Fusion. In *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by Maria Solé, Daniel Recasens and J. Romero, 2609–2612. Barcelona: Futurgraphic.

Inverson, Paul, Lynne Bernstein and Edward T. Auer Jr. 1998. Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. *Speech Communication* 26: 45–63.

Kaye, Jonathan. 1992. Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics* 2: 293–313.

Kiparsky Paul. 2003. Syllables and Moras in Arabic. In *The syllable in optimality theory*, edited by Caroline Féry and Ruben van de Vijver, 147–182. Cambridge: Cambridge University Press.

Kolinsky, Regine and Jose Morais. 1993. Intermediate representations in spoken word recognition: a cross-linguistic study of word illusions. In *EUROSPEECH'93*, 731–734. http://www.isca-speech.org/archive/eurospeech_1993

Liberman, Alvin, Franklin Cooper, Donald Shankweiler and Michael Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review*. 74(6): 431–461.

MacDonald, John and Harry McGurk. 1978. Visual influences on speech perception processes. *Perception and Psychophysics* 24: 253–257.

Mattys, Sven L., and James F. Melhorn. 2005. How do syllables contribute to the perception of spoken English? Insight from the migration paradigm. *Language and Speech* 48: 223–253.

McGurk, Harry and John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264: 746–748.

Öhrström, Niklas and Hartmut Traunmüller. 2004. Audiovisual perception of Swedish vowels with and without conflicting cues. In *Proceedings*, *FONETIK 2004*, 40–43. Stockholm University.

Pan, Ning and William Snyder. 2004. Acquisition of /s/-initial clusters: A parametric approach. In *Proceedings of the 28th Boston University Conference on Language Development*, edited by Alenja Brugos, Linnea Micciulla and Christine E. Smith, 436–446. Somerville, MA: Cascadilla Press.

Prunet, Jean-François, Renée Beland and Ali Idrissi. 2000. The mental representation of Semitic words, *Linguistic Inquiry* 31: 609–648.

Rosenhouse, Judith. 2007. Arabic phonetics in the beginning of the third millennium. In *Proceedings of the16th International Congress of Phonetic Science*, 131–134. http://www.icphs2007.de/

Sams, Mikko, Petri Manninen, Veikko Surakka, Pia Helin and Riitta Kättö. 1998. McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context. *Speech Communication* 26: 75–87.

Segui, Juan and Ludovic Ferrand. 2002. The role of syllabic units in speech perception and production. In *Phonetics, Phonology and Cognition*, edited by Jacques Durand and Bernard Laks, 151–167. Oxford: Oxford University Press.

Sumby, W. H. and Irwin Pollack. 1954. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* 26: 212–215.

Tiippana, Kaisa, Mikko Sams and Riikka Möttönen. 2002. The effect of sound intensity and noise level on audiovisual speech perception. Multisensory Research Conference, abstracts, 36.

Treiman, Rebecca. 1983. The structure of spoken syllables: Evidence from novel word games. *Cognition* 15: 49–74.

——. 1985. Onsets and rimes as units of spoken syllables: evidence from children. *Journal of Experimental Child Psychology* 39: 161–181.
van Wassenhove, Virginie, Ken Grant and David Poeppel. 2007. Temporal window of integration in bimodal speech. *Neuropsychologia* 45: 598–607.
Watson, Janet. 2002. *Phonology and morphology in Arabic*. Oxford: Oxford University Press.
Windmann, Sabine. 2004. Effects of sentence context and expectation on the McGurk illusion. *Journal of Memory and Language* 50: 212–230.
Wright, Richard. 2001. Perceptual cues in contrast maintenance. In *The role of speech perception in phonology*, edited by Elizabeth V. Hume and Keith Johnson, 251–277. London: Academic Press.
——. 2004. A review of perceptual cues and cue robustness. In *Phonetically based phonology*, edited by Bruce Hayes, Robert Kirchner and Donca Steriade, 34–57. Cambridge: Cambridge University Press.

Appendix 1

English word-triples (simple monosyllabic words)

onset	bait—gate—date	/beɪt/—/geɪt/—/deɪt/
	bain[1]—gain—dane[1]	/beɪn/—/geɪn/—/deɪn/
	bad—gad[2]—dad	/bæd/—/gæd/—/dæd/
	bold—gold—doled	/bəʊld/—/gəʊld/—/dəʊld/
	pat—cat—tat[3]	/pæt/—/kæt/—/tæt/
	pill—kill—till	/pɪl/—/kɪl/—/tɪl/
	pod—cod—tod	/pɒd/- /kɒd/—/tɒd/
	[1] are names	
	[2] a form of a steel bar	
	[3] used in conjunction, for e.g. 'tit for tat'	
coda	cheep—cheek—cheat	/tʃɪːp/—/tʃɪːk/—/tʃɪːt/
	bap—back—bat	/bæp/—/bæk/—/bæt/
	kip—kick—kit	/kɪp/—/kɪk/—/kɪt/
	lop—lock—lot	/lɒp/—/lɒk/—/lɒt/
	flap—flack—flat	/flæp/—/flæk/—/flæt/
	map—mack—mat	/mæp/—/mæk/—/mæt/
	tap—tack—tat	/tæp/—/tæk/—/tæt/

English word-triples (branching monosyllabic words)

Onset		
Cr	brill—grill—drill	/brɪl/—/grɪl/—/drɪl/
	brain—grain—drain	/breɪn/—/greɪn/—/dreɪn/
	brew—grew—drew	/bruː/—/gruː/—/druː/
	braze—graze—drays	/breɪz/—/greɪz/—/dreɪz/
	prude—crude—trued	/pruːd/—/kruːd/—/truːd/
	pride—cried—tried	/praɪd/—/kraɪd/—/traɪd/
	prays—craze—trays	/preɪz/—/kreɪz/—/treɪz/
	press—cress—tress	/pres/—/kres/—/tres/
sC	spares—scares—stares	/speəz/—/skeəz/—/steəz/
	spate—skate—state	/speɪt/—/skeɪt/—/steɪt/
	spill—skill—still	/spɪl/—/skɪl/—/stɪl/
	spore—score—store	/spɔː /—/skɔː/—/stɔː/
	spud—scud—stud	/spʌd/—/skʌd/—/stʌd/
	spar—scar—star	/spɑː/—/skɑː/—/stɑː/
	spear—skier—steer	/spɪə/—/skɪə/—/stɪə/
	spool—school—stool	/spuːl/—/skuːl/—/stuːl/
Coda		
Cs	cobs—cogs—cods	/kɒbz/—/kɒgz/—/kɒdz/
	tabs—tags—tads	/tæbz/—/tægz/—/tædz/
	bubs—bugs—buds	/bʌbz/—/bʌgz/—/bʌdz/

	cops—cocks—cots	/kɔps/—/kɔks/—/kɔts/
	tips—ticks—tits	/tɪps/—/tɪks/—/tɪts/
	pups—pucks—puts	/pʌps/—/pʌks/—/pʌts/
	peps—pecks—pets	/peps/—/peks/—/pets/
	maps—macks—mats	/mæps/—/mæks/—/mæts/
cC	harp—hark—hart*	/haːrp/—/haːrk/—/haːrt/
	corp—cork—cork*	/kɔːrp/—/kɔːrk/—/kɔːrt/
	wisp—whisk—whist	/wɪsp/—/wɪsk/—/wɪst/

* articulated rhotically by a Scottish speaker

Arabic word-triples

C-initial	دلَّ—قلَّ—بلَّ	/balla/—/qalla/—/dalla/
		‘wet’—‘few’—‘show’
	دَالْ—قَالْ—بَالْ	/baal/—/qaal/—/daal/
		‘urinate’—‘say’—‘rotate’
C-medial	رَدَعَ—رَقَعَ—رَبَعَ	/rabaʕa/—/raqaʕa/—/radaʕa/
		‘gallop’—‘patch’—‘keep’
	حَداَ—حَكَىَ—حَبَا	/ħabaa/—/ħakaa/—/ħadaa/
		‘creep’—‘report’—‘urge’
	عَـدْلٌ—عَقْلٌ—عَبْلٌ	/ʕablun/—/ʕaqlun/—/ʕadlun/
		‘plump’—‘mind’—‘virtue’
C-final	نَهَدْ—نَهَقْ—نَهَبْ	/nahab/—/nahaq/—/nahad/
		‘rob’—‘donkey’—‘injection’
	أرَادْ—أرَاقْ—أرَابْ	/ʔaraab/—/ʔaraaq/—/ʔaraad/
		‘curdle’—‘pour’—‘wanted’