Probing the Syllabic Structure of Words using the Audio-Visual McGurk Effect

Azra N. Ali (a.n.ali@hud.ac.uk)
Department of Multimedia and Information Systems, Queensgate
Huddersfield, HD1 3DH UK

Michael Ingleby (m.ingleby@hud.ac.uk)
Department of Computing and Mathematics, Queensgate
Huddersfield, HD1 3DH UK

Abstract

Many studies have shown that a syllable has an internal hierarchical structure and is made up of two main constituents: the consonantal onset and the rhyme. Most experiments to test the cognitive reality of the syllable in the mental models of humans have involved word games using words with concealed parts. Here we outline an alternative way of testing for syllabic structure using McGurk fusion. The experiments develop a method for locating syllable boundaries in polysyllabic words, based on the coda-onset fusion rate differences found in our earlier work with monosyllabic words. For the sake of simplicity, the fusion rate measurements are made on polysyllabic words in which an internal consonantal site is given audio-visual incongruence. A word-internal site may be the coda of a first syllable or the onset of a second syllable. The aim of the research was to determine whether a fusion in the word-medial consonant of polysyllabic words confirms the traditional or another syllabification scheme. The results from our first study show that the McGurk fusion rates can be used to locate syllable boundaries in polysyllabic words. The findings show that in 65% of the English words investigated the coda of the traditional morphological stem has the fusion-rate behaviour of an empirical coda.

Keywords: Linguistics; syllabic structure; speech perception; audiovisual speech; McGurk effect.

Syllable Structure using Word Games

"Although nearly everyone can identify syllables, almost nobody can define them…It is curiously difficult to state an objective phonetic procedure for locating the number of syllables in a word or phrase in any language" (Ladefoged, 1993). Most researchers agree that a syllable is more than a string of unrelated phonemes. Many studies make a case for an internal hierarchical structure of an onset and rhyme constituents, the latter being a vowel nucleus and optional coda, (Fudge, 1969; MacKay, 1972; and Ewen and van der Hulst, 2001). Examples are shown in Figure 1 with branching and non-branching coda constituent.

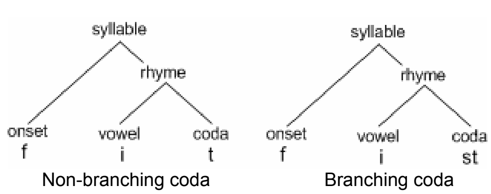

Figure 1: Syllabic structure.

Empirically, the onset-rhyme structure of the syllable has been revealed in several experiments using word-games. The simplest games used monosyllabic words C(C)VC(C) presented as auditory stimuli (Trieman, 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 using perceived inner boundaries, usually located between onset and rhyme. 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 the participants made onset-rhyme partitioning more often than by chance – suggesting the empirical reality of the boundary in the cognitive word model of participants.

In further polysyllabic games using auditory priming stimuli and visual targets, Mehler *et al* (1981) investigated a bimodal monitoring task put to French participants. The participants were presented with an auditory word stimulus for a short duration and the target was a short-duration visual presentation of the first fragment of a text word. The reaction times for identifying the fragment were found to be much shorter when the target corresponded to the whole first syllable of the priming stimulus than when the target was a part syllable. For example, French speakers identified the target syllable /pa/ much faster when primed by *'pa.lace'* than by '*palm.ier'*. Mehler *et al* concluded that word-recognition is dependent upon good placement of masking boundaries at syllabic breaks in words. (These empirical boundaries follow traditional syllable boundaries, at least with francophone participants). In a separate study, Segui (1984), reported similar findings and added that the first syllable can be considered as important key to lexical access. However, in some experiments with English speakers, investigators found no difference between responses after priming by whole or part syllables (Cutler *et al*, 1986). For example, English participants primed by 'balance' and 'balcony' responded similarly to the targets with 'ba-' and 'bal-'. A possible explanation is ambisyllabicity in English, both 'bal.ance' and 'ba.lance' being acceptable syllabifications of 'balance'. Corresponding findings by Zwitserlood *et al* (1993) on English and Dutch participants in priming and masking experiments were similar. Dutch participants given '*bu(k)en'* (= to stoop), *'ro(k)en'* (= to smoke), '*me(n)en'* (= men), showed marginal preference for the coda option in the case of consonants ‘k’ and ‘n’. Another form of bimodal experiment using a picture and word naming task can be found in Ferrand and Segui (1996). The French participants were given an auditory priming stimulus and picture targets. Pairs of picture targets were presented, some where the first syllable of the picture name matched that of the auditory priming stimulus, and others where the first syllable did not exactly match. The matched visual targets elicited a speedier response, more so, with CVC primes than CV.

Variant priming and masking experiments have been performed – all using variation of decision time as a syllabic boundary indicator. These experiments are designed on the supposition that perception of the auditory prime and the separate perception of a visual target (text or image) are mediated by the same structural model of lexical items. It is possible that this identity of mental models is not universal. Therefore we have sought for an empirical probe that engages only one kind of processing, and settled on fused audiovisual stimuli. Thus, the aim of this research is to outline an alternative method for probing perceived syllable structure, using only the response of group of participants to incongruent audiovisual data.

The McGurk Effect

Although speech perception is usually considered as a response to auditory data, the movements of a talker's mouth and face strongly influence what an observer perceives, even when the auditory signal is clear and unambiguous. Evidence for strong interaction between audio and visual speech channels in human speech perception is found in the well-known McGurk fusion effect (McGurk and MacDonald, 1976). If humans are presented with temporally aligned but conflicting audio and visual stimuli – now known as ‘incongruent stimuli’ - the perceived sound may differ from that present in either channel. McGurk and MacDonald asked their recording technician to create a videotape with the audio syllable 'ba' dubbed onto a visual 'ga', most normal adults reported hearing 'da' or 'tha'. But when the participants were presented with only one modality (visual or audio, not too noisy), or stimuli without audiovisual incongruity, they reported the syllables correctly.

The McGurk fusion has been studied intermittently over the last three decades mainly from a psychophysical point of view – using audiovisually incongruent nonsense syllables (e.g. 'ba', 'ga', 'da'; 'aba', 'aga', 'ada' , 'ibi', 'igi', 'idi' etc.). In contrast with nonsense syllables, there are contradicting reports regarding fusion effects in the context of natural words. Easton and Basala (1982) reported that English words with incongruity did not induce McGurk fusion; but Dekle *et al* (1992) reported fusion occurring in real words. A third study was carried out, in Finnish, and also reported fusion in real Finnish words (Sams *et al*, 1998), but fusion was concentrated in onset consonant only.

Our earlier series of empirical investigations on incongruent real-word stimuli to show that measurement of McGurk fusion rates can be used to probe the mental lexicon of participants (Ali and Ingleby 2002; Ali, 2003a; Ali, 2003b; Ingleby and Ali, 2003). In these studies incongruities in monosyllabic CVC words were concentrated at a single segmental site: either in an on onset consonant, or a vowel nucleus (short and long vowel) or a coda consonant, Table 1. Fusion was elicited in all cases.

Table 1: Examples of fusion in CVC words.

Segmental site	Audio channel	Visual channel	Fusion
Onset consonant	**p**at	**t**at	**c**at
Vowel	h**i**d	h**o**d	h**ea**d
Coda consonant	ma**p**	ma**t**	ma**ck**

The first objective was to determine whether the qualitative patterns of response in vowel quality and place of articulation incongruity were the same in the three linguistic segmental sites (onset, nucleus and coda). The studies revealed that fusion was not sensitive to the constituent position: incongruence in coda consonants elicits the same fusion responses as in onset consonants. Quantitatively, however, there is a rate difference: fusion rates for coda consonants are significantly higher than for onsets. Similarly, for vowel incongruity: short vowels were more vulnerable to fusion than long vowels. The observed effects of incongruity site reported in the experiments were not attributable to poor audio or visual signal quality. This was demonstrated in a separate unimodal experiments (audio only and visual only) and in congruent audiovisual experiments.

Testing of Syllabification Hypothesis using Incongruent Audiovisual Data

In our earlier studies with monosyllabic words (Ali 2003a; Ali 2003b; Ingleby and Ali 2003), the significant difference in fusion rates between onsets and codas noted in non-branching words were found to survive in both branches of a branching syllabic constituent. The significant rate differences between the first branch of an onset constituent and the second branch of a coda constituent in branching constituents, (between C_1 and C4, Figure 2) are of particular interest for locating boundaries in polysyllabic words.

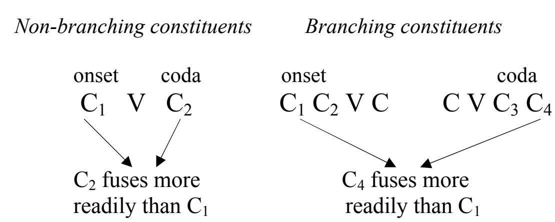

Figure 2: Significant difference in fusion rates.

The coda-onset fusion rate differences survive in polysyllabic words, and therefore can be used to measure their syllabic structure by introducing incongruence at possible boundary sites. The aim is to determine whether

participants treat a word-medial consonant as the coda of a preceding syllable or the onset of a succeeding syllable, Figure 3. Thus, fusion-rate measurements on such consonants allow us to determine whether fusion in the word medial consonant (the second consonant) of polysyllabic words supports a traditional syllabification hypothesis or an alternative, as exemplified in Figure 4.
According to Longman Pronunciation Dictionary – LPD (Wells, 2000), the syllable structure for the word 'coating' is 'coat . ing' (/kəut.ɪŋ/), the consonant /t/ serving as the coda of the first syllable. In language teaching, tradition and morphology decide the location of syllable boundaries.

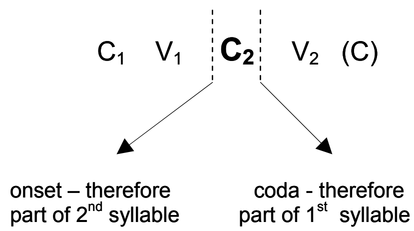

Figure 3: Syllable structure of interest.

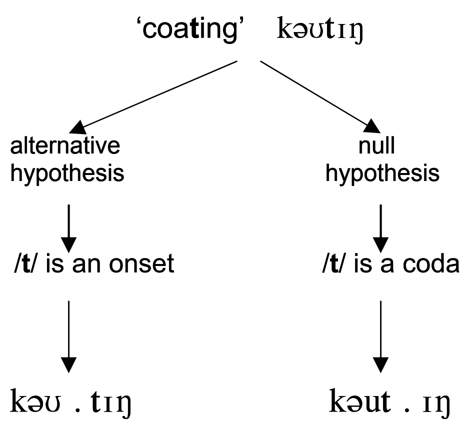

Figure 4: Syllabification hypotheses.

The first experiments (Ali 2003; Ingleby and Ali 2003; Ali 2003a, Ali 2003b) established decision thresholds on fusion rates for consonants in branching and non-branching constituents. In the case of non-branching constituents C_1VC_2…:

fusion rate > 50% → C_2 is a coda (C) (1)
fusion rate < 40% → C_2 is an onset (O) (2)
fusion rate > 40%<50% → C_2 is ambisyllabic (A) (3)

Table 2, shows some typical examples that could be used in experiments to test English syllabification using the McGurk effect. From the threshold fusion rate one can then determine whether the word-medial (second consonant) in these words is behaving like an onset or coda.

Table 2: Test materials for polysyllabic words.

	Incongruent stimuli Audio	Visual	Syllable structure if fusion rate: < 40%	> 50%
1	co**p**ing /kəʊpɪŋ/	co**ck**ing /kəʊkɪŋ/	coa **. t**ing /kəʊ. tɪŋ/	coa**t .** ing /kəʊt .ɪŋ/
2	chee**p**ing /tʃi:pɪŋ/	chee**k**ing /tʃi:kɪŋ/	chea**. t**ing /tʃi:. tɪŋ/	chea**t.** ing /tʃi:t. ɪŋ/
3	bobby /bɒbɪ/	boggy /bɒgɪ/	bo**dd**. y /bɒdd. ɪ/	bo**. dd**y /bɒ. ddɪ/

The aim of the experiment was to cover a wide sample of words, using both low and high frequency words, using, bisyllables (cases 1 and 2 in Table 2) and trisyllables. Also, examples likely to be affected by ambisyllabicity (case 3 of Table 2) were investigated. The feasible decision thresholds proposed above allow confirmation of either a coda hypothesis to the incongruent consonant, or an onset hypothesis, or neither (ambisyllabic case).

Experimental Design

Creating the Stimuli

The English polysyllabic words (meaningful words) used are shown in Table 2. The video recordings were done inside a quiet, controlled laboratory using a standard 8 mm digital (Sony) Camcorder with built-in microphone for audio. From the resultant video recording, each word uttered was captured (both the visual and audio channels) into *.avi files. For the creation of incongruent stimuli, standard editing software (Adobe Premier 5.5) was used. The audio channel from the second member of the pair was imported and aligned with the audio channel of the first member of the pair. After coarse alignment, the first member's original audio channel was erased. The experimenter made fine judgments of proper alignment manually, after previewing the video clip. Incongruent stimuli and a few natural, fully congruent controls were then saved as *.avi files with a frame rate of 25 and a frame size of 640 x 512 pixels.

Procedure

Twenty-four participants took part in the experiment, native British English speakers, 12 females and 12 males, with an age range between 23 to 54 years and no specialized linguistic knowledge. The participants were provided with report forms on which to record 'what they thought the speaker was saying' when receiving an experimental audiovisual stimulus. The report forms included text-words corresponding to the audio channel of the stimulus, to the video channel of the stimulus, to a number of possible results of channel fusion, and some random words. A space to write in a word not included on the form was also provided.

Results

First the data for congruent stimuli were analysed, revealing that 100% of the participants accurately perceived what the speaker was saying. The subsequent aim of the research was to determine whether fusion in a word-medial consonant (usually the second consonant) of polysyllabic words fits traditional syllabification or another model. The results highlight that in 65% of the cases, the traditional coda of the stem is coming out as a perceived coda. Fusion rates for each stimulus are shown below, Table 3.

Table 3 Assigning Syllable Boundary.

	Audio	Visual	Expected & Perceived Fusion	Fusion Rate %	Middle consonant is:
-ɪŋ	bibbing /bɪbɪŋ/ (1)	bidding /bɪgɪŋ/ (51)	Biggin /bɪdɪŋ/ (638)	100.0	C
	tabbing /tæbɪŋ/ (1)	tagging /tægɪŋ/ (95)	tadding /tædɪŋ/ (4)	87.5	C
	knobbing /nɒbɪŋ/ (1)	noggin /nɒgɪŋ/ (51)	nodding /nɒdɪŋ/ (508)	87.5	C
	cheeping /tʃi:pɪŋ/ (7)	cheeking /tʃi:kɪŋ/ (4)	cheating /tʃi:tɪŋ/ (381)	75.0	C
	ribbing /rɪbɪŋ/ (47)	rigging /rɪgɪŋ/ (180)	ridding /rɪdɪŋ/ (47)	75.0	C
	raping /reɪpɪŋ/ (145)	raking /reɪkɪŋ/ (136)	rating /reɪtɪŋ/ (1152)	71.4	C
	whipping /wɪpɪŋ/ (209)	wicking /wɪkɪŋ/ (5)	witting /wɪtɪŋ/ (1)	71.4	C
	clopping /klɒpɪŋ/ (4)	clocking /klɒkɪŋ/ (83)	clotting /klɒtɪŋ/ (45)	62.5	C
	sleeping /sli:pɪŋ/ (2460)	sleeking /sli:kɪŋ/ (3)	sleeting /sli:tɪŋ/ (7)	42.9	A
	coping /kəʊpɪŋ/ (908)	coking /kəʊkɪŋ/ (21)	coating /kəʊtɪŋ/ (355)	0.0	O
-ə	whipper /wɪpə/ (3)	wicker /wɪkə/ (210)	witter /wɪtə/ (19)	87.5	C
	flipper /flɪpə/ (27)	flicker /flɪkə/ (397)	flitter /flɪtə/ (3)	62.5	C
	rubber /rʌbə/ (1555)	rudder /rʌgə/ (34)	rugger /rʌdə/ (292)	37.5	O
	slipper /slɪpə/ (120)	slicker /slɪkə/ (27)	slitter /slɪtə/ (1)	12.5	O
	bibber /bɪbə/ (1)	bigger /bɪgə/ (3488)	bidder /bɪdə/ (412)	12.5	O
-ɪ	cabby /kæbɪ/ (31)	khaki /kækɪ/ (166)	caddie /kædɪ/ (237)	100.0	C
	babby /bæbɪ/ (26)	baggy /bægɪ/ (234)	baddy /bædɪ/ (4)	75.0	C
	dippy /dɪpɪ/ (27)	dicky /dɪkɪ/ (8)	ditty /dɪtɪ/ (41)	54.1	C
	hoppy /hɒpɪ/ (18)	hockey /hɒkɪ/ (597)	hottie /hɒtɪ/ (2)	37.5	O
	Bobby /bɒbɪ/ (1075)	boggy /bɒgɪ/ (70)	body /bɒdɪ/ (24567)	12.5	O

O = Onset C = Coda A = Ambisyllabic
Figures in brackets, ( ), denote word frequency count in the BNC

It appears, however, that the consonant that would be traditionally syllabified as a stem coda tends to behave as an onset before an '–er' affix (-ə is non-rhotic English dialects). This affix lowers the fusion rate of the traditional stem coda to a rate more typical of an onset. Other morphological affixes '–ing' and '–y/ie' do not lower fusion rates of the stem coda to such a large extent. This leads towards the idea that fusion reveals more than one type of morphological affixing, as discussed later in this section.

Vowel context effects The consonant fusion rates were not due to the vowel context in the stem of the word. Fusion rates were analysed in relation to the type of vowel quality, for example, vowel ɪ and iː were regarded as palatal vowels, æ, ɒ and ʌ as velar vowels. Other vowels used in the incongruent stimuli were diphthong vowels. Fusion rates were as follows: palatal vowels 58.6%; velar vowels 62.0% and diphthong vowels 45.5%. Although exploratory analysis indicated that fusion seemed greater for vowels with velar quality, detailed statistical analysis shows that fusion rate is not significantly dependent on the vowel quality ($\chi^2 = 1.83$, df = 1, α = 5% significant level threshold, $p = 0.401 > \alpha$). Therefore, the fusion rates reported in Table 3 are not dependent on the vowel context.

Word salience effects "Words that occur more often in printed language are easier to recognize than less frequently occurring words. This effect is known as the word frequency effect". (Grainger, 1990:228). Studies have shown that participants were quicker to respond to high frequency words than to low frequency words in lexical decision tasks. This is a special case of the effect of salience on perception, where word frequency measures salience. We used this measure of salience, calling on Kilgarriff's (1997) lemmatised list (of words sorted by headword or stem). The list is based on the British National Corpus (BNC), and the frequency of the most common words in the list is more than.

The consonant fusion-rate dependencies on salience are included in Table 2. A perceived fusion response reported by the participants was considered as a high frequency word if it had a frequency count of 800 or more in the BNC. The table shows that some high-frequency perceived words like 'body' (/bɒdɪ/) elicited little fusion (12.5%), whilst some low-frequency fusion words like 'caddie' (/kædɪ/) elicited high fusion rate (100%).

However, word frequency changes when stem words and words with stem and morphological affixes are counted separately. Thus, the classification of responses as common or rare by whole word was not identical to the classification by stem. The results of exploratory analysis indicate, nevertheless, that fusion rates seemed greater for low frequency words, whether stem or whole-word frequencies were used to define salience classes. In fact, a detailed statistical test revealed no significant salience differences: fusion rates were the same for high and low frequency words. Classifying by salience of word gave ($\chi^2 = 2.32$, df = 1, $p = 0.127 > \alpha$) while classifying by stem gave ($\chi^2 = 1.15$, df = 1, $p = 0.283 > \alpha$). There are no significant salience effects amongst fusion rates !

Morphological affixes Although salience effects were ruled out, there appears to be differences in fusion rates depending on the morphological affix. The exploratory analysis shows fusion rates were as follows: 69% before affix '–ing', 63% before affix '–y/ie' and 43% before affix '–er'. Detailed statistical analysis revealed that fusion rate is dependent on the type of affix ($\chi^2 = 14.95$, df = 1, $p = 0.020 < \alpha$). The finding suggests that different morphological affixes can bind differently to a stem, but this needs investigating further in a separate study. There is already evidence from stress patterns for two types of affix: when the word 'parent' is affixed with '-hood' it retains stress on the first vowel, but on affixing with '-al' the stress shifts to the second vowel. Also certain affixes are more prone to induce stem perception errors than others (Janssen and Humphreys, 2002). It has been suggested that a "that morphological information is represented in the mental lexicon in a quite detailed way" (McQueen and Cutler, 1998:424).

Discussion and Conclusion

Experiments to test the cognitive reality of the syllable in the mental models of humans have mainly involved word games using words with concealed parts. In this paper an alternative way of probing the syllabic structure was tested using the coda-onset fusion rate thresholds to locate syllable boundaries in polysyllabic words. The aim of the research was to determine whether fusion-rates in the word medial consonants (the second consonant) of polysyllabic words support traditional or another syllabification scheme.

Firstly, the experiment showed that the coda-onset fusion differences remain in polysyllabic words. The consonant fusion rates were not attributable to salience effects. Secondly, fusion rates reported were not dependent on the vowel quality either. Finally, the results indicate that in most cases, the coda of the stem is coming out as an empirical coda, especially with the syllabified as a stem coda tends to behave as an onset with '–er' affix (-ə is non-rhotic English dialects). Other morphological affixes '–ing' and '–y/ie' do not lower the coda like fusion rates of the stem coda.. It appears that certain morphological affixes like '–y/ie' ('hoppy', 'body' etc.) and '–ing' ('whipping', clocking', etc.) elicit more fusion in the foregoing stem than the agent affix '–er' ('slipper,', 'bidder', etc.). Perhaps the fusion results are indicating that the lexemes and morphemes are two separate processing elements. This needs to be investigated further in a separate study.

Coda-Onset Differences and CV Languages

The abiding pattern, in all our work cited in this paper, is that fusion rates remain significantly different for codas and onsets. This difference would not be observed in pure CV languages, also known as 'no coda languages. In the case of languages for which syllabification is contested, one could use incongruent stimuli to seek for the segments that show too much fusion to be onsets, thereby putting linguistic conjectures and hypotheses to empirical test.

The Arabic tradition of Sybawaih, on which the (phonetic) Arabic alphabet is founded, uses CV units symbolised orthographically by a consonant with a vowel diacritic. The Western tradition of classical scholars treats Arabic like Latin and Greek, postulating that there are CVC, CVVC, CVCC syllables. Recently a number of linguists have argued for a neo-Sybawaihan representation of Arabic speech patterns (see, for example, Baothman (2002) for a review of the arguments).

We have started new research to test whether native Arabic speakers have codas in their mental models of Arabic speech (Idrissi, Ali and Ingleby, 2004). At this early stage of the research, there is a need to check that the fusion patterns for Arabic are qualitatively similar to those of English, Finnish and the languages where McGurk fusion is independently attested by several researchers. We have confirmed that Arabic speakers experience fusion in response to incongruent stimuli. In onset position, for example, the audio stimulus /bæːl/ (= awareness), aligned with /qæːl/ (= chit-chat) in the visual channel is perceived as /dæːl/ (= Arabic letter equivalent to English 'd'). The fusion rates amongst speakers of Egyptian, Saudi and Jordanian dialects are not very different from those we report for English incongruent onsets in this paper. The design of experiments has to involve words syllabified both in the classical tradition with codas and the neo-Sybawaihan tradition without codas, and this, with fusion rate results, will be the subject of another paper.

References

Ali, A.N., &Ingleby, M. (2002) Perception difficulties and errors in multimodal speech: The case of vowels. *Proceedings of the 9th Australian International Conference in Speech Science and Technology,* (pp. 438-444).

Ali, A. N. (2003a). Perception difficulties and errors in multimodal speech: The case of consonants. *Proceedings of the 15th International Congress of Phonetic Sciences*, (pp. 2317-2320). Recasens, D & Romero, J (eds).

Ali, A.N. (2003b). *Speech perception of branching constituents: Experimental comparison of McGurk fusion rates*, (Research Report Series RR0303), School of Computing and Engineering, University of Huddersfield.

Baothman, F. (2002). *Phonology-based automatic speech recognition for Arabic.* Unpublished doctoral dissertation, School of Computing and Engineering, University of Huddersfield.

Cutler, A., Mehler., Norris, D & Segui, J. (1986). The syllable's differing role in the segmentation of French and English', *Journal of Memory and Language*, *25*, 385-400.

Dekle, D. J., Fowler, C. A., & Funnell, M. G. (1992). Audiovisual integration of real words. *Perception & Psychophysics*, *51*, 355–362.

Easton, R.D., & Basala, M. (1982). Perceptual dominance during lip-reading. *Perception & Psychophysics*, *32*(6), 562-570.

Ewen, C.J., & van der Hulst, H (2001). *The phonological structure of words*, Cambridge: Cambridge University Press.

Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, *35*, 708-723.

Fudge, E.C. (1969). Syllables. *Journal of Linguistics*, *5*, 253-286.

Grainger, J. (1990). Word frequency and neighbourhood frequency effects in lexical decision and naming', *Journal of memory and language*, *29*, 228-244.

Idrissi, A., Ali, A.N. & Ingleby, M. (2004). *Incongruent segments and McGurk fusion in Arabic words* (Research Report), University of Huddersfield.

Ingleby, M., & Ali, A.N. (2003). Phonological primes and McGurk fusion. *Proceedings of the 15th International Congress of Phonetic Sciences*, (pp. 2609-2612).

Janssen, D. & Humphreys, K. (2002). Morphological speech errors on agentive and comparative affixes. *Poster Presented at the Third International Conference on the Mental Lexicon*. Banff, Canada.

Kilgarriff, A. (1997). Putting Frequencies in the Dictionary, *International Journal of Lexicography, 10* (2), 135-155.

Ladefoged, P. (1993). *A course in phonetics*. New York: Harcourt Brace Jovanorich.

MacKay, D.G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, *3*, 210-227.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

McQueen, J.M., & Cutler, A. (1998). Morphology in word recognition. In A. Spencer & A.M. Zwicky (Eds.), *The handbook of morphology*, Oxford: Blackwell Publishers.

Mehler, J., Dommergues, J.Y., Frauenfelder, U.H., & Segui, J. (1981) The syllable's role in speech segmentation, *Journal of Verbal Learning and Verbal Behaviour*, *20*, 298-305.

Sams, M., Manninen, P., Surakka, V., Helin, P., & Kättö, R. (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, J. (1984). The syllable: A basic perceptual unit in speech processing. In H Bouma, H and D.G. Bouwhuis, (Eds) *Attention and performance X: Control of language processes*, Lawrence Erlbaum Associates.

Trieman, R. (1983). The structure of spoken syllables: Evidence from novel word games, *Cognition*, *15*, 49-74.

Treiman, R. (1985) Onsets and rimes as units of spoken syllables: evidence from children, *Journal of Experimental Child Psychology*, *39*, 161-181.

Wells, J.C. (2000) *Longman pronunciation dictionary*. Second Edition, Essex: Pearson Education Limited.

Zwitserlood, P., Schriefers, H., Lahiri, A., & van Donselaar, W. (1993). The role of syllables in the perception of spoken Dutch, *Journal of Experimental Psychology*: *Learning Memory and Cognition*, *19* (2), 260-271.