Gradience in morphological decomposability: Evidence from the perception of audiovisually incongruent speech

AZRA N. ALI and MICHAEL INGLEBY

University of Huddersfield

Abstract

Over the last three decades, priming and masking experiments and corpus frequency studies have dominated attempts to find ranking in the decomposability of words containing morphological affixes. Here we establish feasibility of using another experimental probe based on audiovisually incongruent speech stimuli. In response to such stimuli, a proportion of participants report percepts that differ in place of articulation from either the audio or the visual signal, typically reporting percept /t/ when receiving audio /p/ dubbed onto visual /k/. We study the systematic variation of this proportion, the McGurk fusion rate, using a small corpus with affixes -ing*,* -s*,* -er*, and* -y/-ie *bound to simple English stems. We discuss experimental arrangements designed to ensure that participants are not experiencing confusion due to noisy signals, nor experiencing experimenter effects. Generally we find that embedding in a complex word lowers fusion rate, but the different affixes cause different amounts of lowering. We argue that fusion-rate lowering is a measure of the degree of decomposability, and relate this measure to others. These include measures derived from phonetic stress changes brought about by affixing and measures derived from relative frequencies of stem and affixed derived or inflected words in corpora.*

1. Introduction

Recently, textual corpus frequencies have been used in gradience studies, and they usefully probe morphology in terms of the conditional probabilities of associating affixes with a stem (Hay 2003; Hay and Plag 2004). Using this measure Hay (2003) demonstrates that affixes bind with a range of strengths to stems and other affixes, indicating that there is ranking in the decomposability of morphologically complex words. In lexical phonology, this ranking has been concerned more with the phonological effects of decomposability than with frequencies, and has led to attempts to rank the degree of decomposability of words with affixes (Dixon 1977, cited in Booij 2005, and other references cited therein). A good example of

JLP 1 (2010), 263–282
DOI 10.1515/LABPHON.2010.013
1836-6346/10/01–263

differential decomposability is the effect on the stem *parent-* of derivational affixes *-hood* and *-al*: /ˈpeərənthʊːd/ *versus* /pəˈrentəl/. Affix *-hood* binds more weakly to the stem to alter the stress only a little or not at all from /ˈpeərənt/, while affix *-al* binds more strongly to create a new morphological domain, throughout which stress and related vowel qualities are redistributed. The impact of morphological decomposability on phonetics is not only lexical but also post-lexical, crossing word boundaries. This can be seen in cases of stress redistribution in a noun under the influence of a customary adjective. For example, in (1), the stress distribution in the adjective has moved to emphasize the first vowel nucleus instead of the third in the isolated word. Here, noun-adjective association in English is strong enough to bring about redistribution, but most noun-adjective pairings are more decomposable and do not bring about such effects.

(1) *fundamental* + *error* → *fundamental error*
/fʌndəˈmentəl/ + /ˈerə/ → /ˌfʌndəmentəlˈerə/

Post-lexical phonology is concerned with the systematic nature of such redistribution in word-pairs and phrases and is very much language dependent. There is evidently a complex relationship between strength of association in corpora and the degree of decomposability. In this paper, we are concerned principally with morphological decomposability. But rather than looking at stress and vowel harmony as indicators of degree of decomposability, we examine another route – based on the McGurk fusion elicited by audio-visually incongruent speech. Our stimuli are constructed from recordings of morphologically complex words. These are dubbed recordings of natural speech in which new audio tracks are aligned with videos of a speaker pronouncing English words. The alignment is chosen to synchronize audio and video channels, except in one phonetic segment in which the visual and audio channels differ ('are incongruent'). The classic McGurk-MacDonald experiment (MacDonald and McGurk 1978; McGurk and MacDonald 1976) on audiovisual fusion in onset of nonsense syllables is represented in example (2). The example represents the outcome of presenting an audiovisual incongruent speech stimulus with /ba/ in the **A**udio channel and /ga/ in the **V**isual channel to a group of participants who reported a **F**usion percept /da/ in 64% of the cases, while 9% reported **A**udio /ba/ and 27% reported the **V**isual channel /ga/, as illustrated in example (2):

(2) $_{\mathrm{A}}(\mathrm{ba} \,||\, \mathrm{ga})_{\mathrm{V}} \rightarrow (\mathrm{da})_{\mathrm{F}\,(0.64)}, (\mathrm{ba})_{\mathrm{A}\,(0.09)}, (\mathrm{ga})_{\mathrm{V}\,(0.27)}$

1.1. *Incongruent speech data*

There is a long history of using incongruent speech in laboratory 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. It was called phonological fusion by the pioneer investigator Cutting (1975) who conducted a wide variety of experiments using speech sounds modified to probe the cognitive processes of speech perception. Cutting used audio stimuli modified at a single segmental site so that the synchronous left-ear (L) and right-ear (R) signals differed, and he characterized responses that differed from either signal as the outcome of phonological fusion. A typical outcome of dichotic phonological fusion can also be represented using our succinct form:

(3) $_{\text{L}}$(ba: || ga:)$_{\text{R}}$ → (da:)$_{\text{AF (0.56)}}$, (ba:)$_{\text{L (0.09)}}$, (ga:)$_{\text{R (0.35)}}$

The representation simply means that two synchronized signals are presented in parallel (||) to left and right ears, produced responses at the stated rates. The percepts elicited by the stimulus were as follows: 56% of the group reporting the **A**coustic **F**usion percept /da/, and the remainder reporting either the **L**eft (9%) or the **R**ight (35%) channel signal. It is likely that these specific ratios of listeners' responses could be affected by quality of speech, signal-to-noise (SNR) at the listener's ear. The most important point from the study is that there was a strong /da:/ response and that there was a secondary /ga:/ response. Halwes (1969, cited in Cutting 1976: 121) found that a new percept was perceived more often than by chance even when the dichotic pair had different or same fundamental frequency. The fusion was called 'phonological' because the reported percepts conformed to the phonotactic rules of English, as exemplified by:

(4) $_{\text{L}}$(leɪ || peɪ)$_{\text{R}}$ → (pleɪ)$_{\text{AF}}$, (leɪ)$_{\text{L}}$, (peɪ)$_{\text{R}}$, (lpe)$_{\text{AF}}$
(5) $_{\text{L}}$(tæs || tæk)$_{\text{R}}$ → (tæsk)$_{\text{AF}}$, (tæks)$_{\text{A}}$, (tæs)$_{\text{L}}$, (tæk)$_{\text{R}}$

In these succinct representations, responses are listed in descending order of rate of occurrence. In (4), even when a /leɪ/ signal was temporally ahead of a /peɪ/ in the other ear, Cutting's participants tended to report the phonotactically licit /pleɪ/ at a high rate, and almost none reported the illicit /lpeɪ/. The fusion responses in (5) /tæsk/ or /tæks/ occurred at almost equal rates, as might be expected since both percepts are, of course, licit in English phonotactics. Other types of experiments with laboratory-modified speech involve stimuli such as:

a) masked or gated speech (target word stimuli with initial or final segments cut out, primed by a preceding whole-word stimulus, e.g. Cutler et al. 1986; Segui and Ferrand 2002; Taft 1979 etc.).
b) primed speech (target word stimuli preceded by a related priming signal – masked audio or textual, sometimes pictorial, e.g., Segui and Ferrand (2002) and many others).
c) speech with groups of incongruent segments, so-called migration studies in which the stimuli evoke a migration illusion, exemplified by Kolinsky et al. (1995), and Mattys and Melhorn (2005) in (6) and (7).

(6) $_{\text{L}}$(dɒlmal || kɛ:rfɪn)$_{\text{R}}$ → (dɒlfɪn)$_{\text{M}}$, (kɛ:rmal)$_{\text{R}}$, (dɒlmal)$_{\text{L}}$, . .
(7) $_{\text{L}}$(bi:tɔ || kɔʒʊ:)$_{\text{R}}$ → (bi:ʒʊ:)$_{\text{M}}$, (bi:tɔ)$_{\text{L}}$, (kɔʒʊ:)$_{\text{R}}$

Generally, when priming of types in (a) and (b) above is used in experiments, the speed and accuracy of recognition depends critically on the shared features of prime and target. Therefore, these probes indicate what feature units mediate perception. However, it has been the practice in these types of experiments to use closed-choice response formats, and these have recently been called into question. Closed-choice formats require participants to "make verbal and/or planned overt motor responses, make unnatural metalinguistic decisions, and understand explicitly task-related instructions" (Shaker et al. 2006:26) and such metalinguistic burdens make the interpretation of measurements problematic.

In the migration example (6) the response rates were in fact that 50% of participants reported the migration percept /dɒlfɪn/, 20% reported the migration percept /kɛ:rmal/, 15% reported the left channel signal /dɒlmal/ and 15% reported the right channel signal /kɛ:rfɪn/. The migration responses are percepts that recombine contiguous groups of segments from different channels of the incongruent stimulus to create something 'illusory', as if material from left and right channels were migrating into a single percept channel. This kind of experiment mobilizes quite different cognitive activity than the McGurk and related effects, since entire pieces are moved around – morpho-syntactic activity. The migration illusion occurs at a high rate when the segment groups from different channels can recombine to form a familiar word in the mental lexicon of participants – in the French example (7), the word is 'bijou', analogous to 'dolphin' in the English example (6). In experiments with Francophone participants (Kolinsky et al. 1995) and with Anglophones (Mattys and Melhorn 2005), the recombinative illusions with the highest rates occur as migrations of syllables (rather than segments or other units) from dichotically-presented polysyllables. Such experiments indicate the empirical presence of units that are close correlatives of the syllables of traditional linguistic analysis. Our use of the illusions elicited in McGurk fusion is analogous to such migration studies. We have used them to probe several organizational features of the mental lexicon for empirical correlates of syllabic onsets and codas and the degree of decomposability of affixes.

1.2. *McGurk fusion and the mental lexicon*

McGurk and MacDonald (1976) pioneered the investigation of audio-visually incongruent dubbed speech sounds to get results represented by:

(8) $_{A}(\text{ba} \,\|\, \text{ga})_{V} \rightarrow (\text{da})_{F(0.64)}, \ldots$ (case of alveolar fusion)
(9) $_{A}(\text{pa} \,\|\, \text{ka})_{V} \rightarrow (\text{ta})_{F(0.50)}, \ldots$ (case of alveolar fusion)

In their first study, all the fusible segments all had the same plosive manner of articulation, but differed in place of articulation and voicing state. The results revealed that when participants were presented with conflicting audio and visual channel, place fusion is greatest with a labial sound in the audio and a velar gesture in the visual stimulus. In a second study, however, they included nasal segments from a different manner class, investigating all possible auditory-visual combinations of eight CV syllables /ba ga pa ka ta da ma na/ (MacDonald and McGurk 1978). The study revealed, *inter alia*, a different kind of phonological fusion across manner classes.

(10) $_{A}$(na || ba)$_{V}$ → (ma)$_{F(0.64)}$, . . . (case of nasal fusion)
(11) $_{A}$(na || pa)$_{V}$ → (ma)$_{F(0.25)}$, . . . (case of nasal fusion)

In such cases, participants appear to pick up the nasal manner of articulation from the audio signal and the labial place of articulation from the visible lip gesture. The percept appears to be the outcome of a cross-channel nasal assimilation process of a distinctly phonological nature.

Working with 'illusory' phenomena in dichotic migration or audio-visual fusion calls for appropriate reality checks. In the case of McGurk fusion, checking methods have been devised in many laboratories over more than thirty years. The illusion is persistent even when participants are told that they are being shown recordings in which the sound-track does not match the visual lip movements of the speaker. It survives size reduction of the video image. It is also robust against acoustic noise (Tiippana et al. 2000; and others) and against visual noise (Fixmer and Hawkins 1998). McGurk fusion phenomena are now known to survive embedding in the many natural languages, occurring amongst speakers of different mother tongues: French (Colin et al. 1998); Dutch (Gelder et al. 1995); Finnish (Sams et al. 1998) and Chinese (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 (Barutchu et al. 2008; Brancazio 2004). Broadly speaking, these studies showed that McGurk fusion rate is sensitive to cues that change a participant's expectation of signal context and is sensitive to lexical and semantic context.

Other studies (illustrated in representations (12) to (18)) have used the McGurk effect and fusion rates to probe syllabic structure in English and Arabic (Ali and Ingleby 2004; Ali et al. 2005). The results (cf. Table 1) showed onset-coda difference for English stimuli: fusion rates were greater in syllable codas than in syllable onsets. For Arabic stimuli with Arabophone participants, no coda-onset difference was observed, which indicates that Arabic is following a CV structure as inferred by evidence on stress and automatic speech recognition (Baothman and Ingleby

Table 1. *Average fusion rates at syllable onset and syllable coda sites*

Stimuli and participants	onsets	codas	t-test*
English Stimuli–Anglophone Participants	40%	65%	Yes
Arabic Stimuli–Arabophone Participants	67%	74%	No
Arabic Stimuli–Anglophone Participants	39%	75%	Yes

* tests whether onset & coda statistical distribution of fusion rates are significantly different (Ali and Ingleby 2004; Ali et al. 2005; Ali et al. 2008)

2002). But when Anglophones were subjected to the same Arabic stimuli (which were nonce words to Anglophones), they showed a coda-onset difference, perceiving similar fusion patterns to their own native language (see Table 1). The speakers used to make these Arabic stimuli were Arabophones who pronounced all their consonants as onsets, so clearly the Anglophone responses resulted from their mental processing.

(12) $_A$(**p**ɪt || **k**ɪt)$_V$ → (**t**ɪt)$_F$, (**p**ɪt)$_A$, . . . (case of incongruent onset)
(13) $_A$(mæ**p** || mæ**t**)$_V$ → (mæ**k**)$_F$, (mæ**p**)$_A$, . . . (case of incongruent coda)
(14) $_A$(h**ɛ**d || h**ʊ**d)$_V$ → (h**æ**d)$_F$, (h**ɛ**d)$_A$, . . . (case of incongruent short vowel)
(15) $_A$(h**ɑ:**d || hi:d)$_V$ → (h**aʊ**d)$_F$, (h**ɑ:**d)$_A$, . . . (case of incongruent long vowel)

(16) $_A$(**b**alla || **q**alla)$_V$ → (**d**alla)$_F$, ... (case of onset incongruence in Arabic)
$_A$(بلّ || قلّ)$_V$ → (دلّ)$_F$, ...

(17) $_A$(ra**b**aʕa || ra**q**aʕa)$_V$ → (ra**d**aʕa)$_F$, . . . (word-medial incongruence in Arabic)
$_A$(رَبَع || رَقَــعَ)$_V$ → (رَدَع)$_F$, ...

(18) $_A$(naha**b** ||, naha**q**)$_V$ → (nahad)$_F$, ... (word-final incongruence in Arabic)
$_A$(نَهَبْ || نَهَقْ)$_V$ → (نَهَدْ)$_F$, ...

The general patterns emerging from this growing body of work on the persistence of fusion percepts are thus:

(a) that fusion phenomena survive even in unfavorable circumstances,
(b) that fusion rate can be enhanced by favorable experimental design,
(c) that fusion responds quantitatively but not qualitatively to any context in which an incongruent segment is embedded.

In this paper, we present evidence that quantitative rates of McGurk fusion are also influenced by morphological factors. This offers the possibility of complementing evidence gathered from corpus frequency and stress redistribution studies of decomposable of words with affixes. The focus of the complementary investigation was to discover whether or not comparative fusion rates can provide an experimental probe of the degree of decomposability of some stem + affix words. As is usual in other laboratory experiments with manipulated speech, it is assumed that the cognitive processes engaged during perception of manipulated speech are essentially the same as those employed in the interpretation of natural speech. To add validity to this ecological assumption, the manipulated data were randomly interspersed with natural speech data. The current study leaves open for future research the investigation of more complex words with stem and affix, as well as the possible correlation of fusion rates with post-lexical structure (phonology across word boundaries) and with semantic cues.

2. Morphological decomposability and fusion rates

In examining the impact of morphological affixes on fusion at stem sites in some English *stem* + *affix* words[1], we used a small sample of inflectional suffixes *-ing, -s* and derivational suffixes such as *-er, -y*/*-ie*, considered to be amongst the most productive in English. Incongruence was sited in syllable coda position of the stem – as in the following case of syllable coda incongruence before inflectional *-s*:

(19) $_{A}$(tæ**b**z || tæ**g**z)$_{V}$ → (tæ**d**z)$_{F}$, . . .

2.1. *Participants*

Forty-four participants took part in the experiment. They were native British English speakers, with an age range between 23 to 54 years. None had specialized knowledge of linguistics or psychology, but were from business and computing disciplines. All participants had normal or corrected-to-normal vision and none reported any hearing problems.

2.2. *Stimuli*

Our word set is exemplified by examples in (20) to (23). Some of the words used as stimuli had a structure with more than one morphological interpretation – "comorphology" of homophones with another morphological structure. In the experimental design, we limited this by use of response sheets to record 'what they thought the speaker was saying' for each experimental stimulus. The response sheets listed options corresponding to the words in the audio and video channels, the expected fusion response, and finally, a space to write in any word not explicit in the list. Many of the stimuli (all real words)[2] had high stem frequencies and low derived and inflected word frequencies in the British National Corpus (BNC). Examples of stimuli, in the 4 suffix groups are shown below, and a full list of the stimuli can be found in Appendix 1.

(20) $_{A}$(reɪpɪŋ || reɪkɪŋ)$_{V}$ → (reɪtɪŋ)$_{F}$ (cases of *stem* + *ing,* gerund)
$_{A}$(tæbɪŋ || tægɪŋ)$_{V}$ → (tædɪŋ)$_{F}$

(21) $_{A}$(mæps || mæks)$_{V}$ → (mæts)$_{F}$ (cases of *stem* + *-s*, plural)
$_{A}$(kɔ:ps || kɔ:ks)$_{V}$ → (kɔ:ts)$_{F}$

(22) $_{A}$(kæbɪ || kækɪ)$_{V}$ → (kædɪ)$_{F}$ (cases of *stem* + *-y*/*-ie*)
$_{A}$(bæbɪ || bægɪ)$_{V}$ → (bædɪ)$_{F}$

(23) $_{A}$(flɪpə || flɪkə)$_{V}$ → (flɪtə)$_{F}$ (cases of *stem* + *-er*, agent)
$_{A}$(slɪpə || slɪkə)$_{V}$ → (slɪtə)$_{F}$

We did not select homophone words like 'rex' / 'wrecks', 'mist' / 'missed' etc., as these would complicate the designing of word-triple stimuli. For experimental purposes we needed the word in the audio channel, the word in the visual channel and an expected fusion word to share a common morphology.

The test stimuli were 30 polymorphemic incongruent audiovisual stimuli, but for the sake of ecological validity, avoidance of experimenter effects etc., 25 polymorphemic congruent audiovisual control stimuli were embedded randomly amongst incongruent stimuli. There were also 20 monosyllabic stimuli to distract participants from the morphology focus of the experiment. The 75 stimuli, i.e. test and filler items, were organized into three groups and participants were given random sequences from only one of the groups.

2.3. *Creating congruent and incongruent stimuli*

Video recordings of the audio word, visual word and the expected fusion word (Appendix 1) were made inside a controlled laboratory using a standard 8 mm digital (Sony) Camcorder with built-in microphone for audio. The speaker uttered each word twice, to enable best selection from the two, in terms of audibility and visibility. The speaker produced the words at a comfortable speaking rate to prevent effects of coarticulation of the sort that might affect place perception. The video recording was spliced into individual word recordings and the average voice onset occurred at 60 ms after video clip onset. These splices were stored in a standard *.avi file format using the 640 × 512 pixel frame size familiar to video watchers; and all stimuli were presented at the customary 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, say /mæ**p**s/ (video 1) with /mæ**k**s/ (video 2). The audio channel from video 1 of the pair was imported into and aligned with the audio and visual channel of video 2 of the pair. The experimenter made fine judgments of proper alignment manually, after previewing the video clip. After alignment, the audio of /mæ**k**s/ (video 2) channel was erased. The resultant video was thus /mæ**p**s/ in the audio channel aligned with visual lip movements saying /mæ**k**s/. All data (congruent and incongruent audio-visual stimuli) were stored in a standard *.avi file format as described above.

2.4. *Instruction sheet and response form and procedure*

Many priming experiments opt for a closed-choice format for reporting percepts, often using “yes” and “no” buttons and asking participants to report as quickly as possible. To minimize the metalinguistic load on participants, we gave participants an open choice when reporting their percepts; and to minimize the risk of guessing experimenter intentions, we adopted the double-blinding[3] technique. The response forms listed the text words corresponding to audio and visual channels, a possible fusion response, two random words with the same morphological structure as the audio and visual originals, and a space to write in any word not present in the list.

2.5. *Procedure*

The experiment was carried out in a controlled laboratory with minimal background noise. Participants sat about half a metre from a 17" monitor screen and used headphones connected to the computer to listen to the audio. The headphones maximized the audio signal and further excluded background noise effects. Dialogue between participants and experimenter was controlled using a standardized instruction sheet. The term 'hearing' was not used on the sheet and participants were asked to wear the headphones provided, watch the video and decide what the speakers were saying by noting their response on the response form. There were no time limits, set or measured, to complete the experiment and participants were allowed to replay the stimuli. Also, no feedback about the experiment was given to them.

2.6. *Results*

Firstly, the data for congruent stimuli which were embedded amongst the incongruent stimuli were analyzed in terms of accuracy, revealing that 100% of the participants accurately perceived what the speaker was saying, eliminating poor vision or hearing as a factor influencing results. Arranging that participants were all at ceiling was a normative control, and excludes the possibility of tracking individual variation in acuities. But such variations were not part of the investigation. Secondly, responses to incongruent stimuli were compiled. The overall proportions of audio channel, visual-channel and real word fusion responses are summarized in Table 2 (response rates to individual stimuli can be found in Appendix 1).

We noted that a small percentage (2.6%) of the responses were nonce words recorded in the open-choice option of the response form and are detailed in Appendix 1. This indicates that participants did not hesitate to report their own response. We included these non-words as fusion responses because responses were dental /θ/ and /ð/, which did not exist in the audio or visual channel. Similar findings were noted in the work of MacDonald and McGurk (1978). When fusion was not reported, participants opted more towards the audio channel. If the audio signal had been weak then we would have seen a bias towards the visual channel, as reported by Clement and Carney (1999: 2271) who showed in an incongruent audio-visual experiment that "the degree of visual bias is in fact related to degree of auditability", i.e. the poorer the audio the more bias towards the visual channel. In the work of MacDonald and McGurk (1978), the average fusion rate in CV syllables was 61% and in our current study with incongruent segment in the syllable coda of

Table 2. *Average response rates for all incongruent stimuli*

Audio	Visual	Fusion
26.0%	15.1%	58.9%

the stem the average fusion rate was 58.9%. However, the fusion rates varied systematically across the four affixes under investigation, as discussed in the following Section 2.6.1 and Section 3.

2.6.1. *Morphological affixes*

The fusion rates across the four suffixes were; *-ing* (69%), *-y*/*-ie* (63%), *-s* (54%) and *-er* (43%). Generally, the effect of affixing lowers the fusion rate considerably in the stem, as exemplified by (24) and (25) below.

(24) $_{A}$(mæp || mæk)$_{V}$ → (mæt)$_{F\ (0.80)}$, (mæp)$_{F\ (0.20)}$
(25) $_{A}$(mæps || mæks)$_{V}$ → (mæts)$_{F\ (0.27)}$, (mæps)$_{F\ (0.64)}$, (mæps)$_{F\ (0.09)}$

Our simple monosyllabic word stimuli (as illustrated in (24)) were part of a separate CVC (real words) experiment. Comparison of fusion rates between stimuli represented in (24) and (25) was not rigorously paired. The two experiments used common equipments, speakers, and procedures, but not all participants were shared. Nevertheless the large reduction in fusion rate is clear[4].

A summary of fusion rates for CVC was given earlier in Table 1, in the line English Stimuli – Anglophone Participants. We can see in example (25) that affixing lowered the fusion rate from 80% in (24) to 27%, a lowering of about 53 percentage points or a percentage lowering of fusion rate of 66%. The average fusion rates at the syllable coda of the stem were not lowered by the same percentage rate by different morphological suffixes as shown in the schematic distribution of the fusion data presented in Figure 1. The distribution curves are made by fitting fusion rates for the four different stems to a Gaussian distribution: the rate is the sum

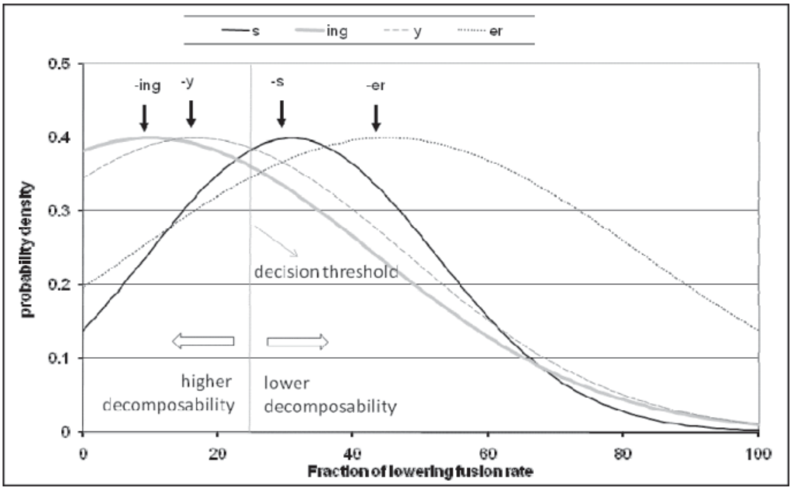

Figure 1. *Schematic distribution of fraction of lowering of fusion rates*

of small contributions from many participants and the central limit theorem predicts an asymptotically Gaussian curve.

The lowering of fusion rates by affixing illustrated in Figure 1 was dependent on the stem, with root mean square variability averaging at 35%. But the mean lowering varied systematically with affix: *-ing* (10.08%), *-y*/*-ie* (16.84%), *-s* (30.09%) and *-er* (45.02%), showing some gradience in morphological decomposability involving the four different morphological suffixes used in this study. The non-parametric Kruskal-Wallis test was used to test the whether this effect was systematic, since participants only viewed a sample of the stimuli and the data collected did not necessarily follow equal distribution laws. The results revealed a main effect for type of suffix ($K = 85.21$, $df = 3$, $p < 0.001$). This parallels results from Hay (2003) and Hay and Plag (2004), whose conditional probabilities demonstrated ranking amongst affixes binding to a range of stems and other affixes. Figure 1 shows similar ranking and gradience, but only four suffixes were tested and the current study does not show where the other English affixes might fall on the same measure of decomposability. This can only be established with an extended study, testing more affixes and stems.

It is now known that relative frequency measures gradience in morphological decomposability (Hay 2003). For example, if a derived word is more frequent than its stem (base) word, then the derived word is likely to be accessed quickly as a whole. If the derived word is less frequent than its stem (base) then it is likely to be accessed in stages, hence less rapidly (Hay 2003). Many of our words had high stem frequency, but low whole word frequency in the British National Corpus (BNC), and thus were likely to be processed as complex units.

To explore further the links between relative frequencies of stimuli and fusion rates at a morphemic boundary, we posited a simple linear regression model. The fusion rate Y_{ij} (dependent variable), under influence of an affix attached to stems of a word-triple j, was taken to be

$$Y_{ij} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + \varepsilon_{ij}$$

in which

X_{1j} is Hay's ratio measure of stem frequency to word frequency for the audio channel
X_{2j} is the corresponding ratio for the visual-channel word
X_{3j} is the corresponding ratio for the fusion word report by the larger number of participants
and ε_{ij} is the statistical residual.

The independent Xs are measures of decomposability used in the BNC (which consists of 100 million words).

For the small sample with only 29 word-triples put to the participants, the regression coefficients for β_0, β_1, β_2, and β_3 were, respectively, 49.82, 0.10, 0.06 and 0.00075. Post-regression analysis of the results was focussed on three hypotheses:

H_{01} – that the decomposability of the audio word has no effect on fusion rate ($t = 8.40$, weakly accepted, $p = 0.053$)
H_{02} – that the decomposability of the visual word has no effect on fusion rate ($t = 0.172$, weakly accepted $p = 0.098$)
H_{03} – that the decomposability of the dominant fusion response has no effect on fusion rate ($t = 0.17$, strongly accepted, $p = 0.868$)

These significance tests indicate that the decomposability of the audio word has at most a marginal effect on fusion rates, and the decomposability of the visual and fusion words have no significant effect.

Early attempts at ranking decomposability in lexical phonology, based on phonetic effects rather than corpus measures or fusion phenomena, have attributed decomposability to affix rather than stem. This attribution (Dixon 1977, cited in Booij 2005) has resulted in a classification of affixes, as reviewed by Kaisse (2005). The classification was approximated to a dichotomy: with *-age, -al, -ant, -ance,* showing a tendency lower decomposability, while *-ary, -ate, -hood, -ic, -ion* etc., showing a tendency higher decomposability. The comparative fusion measures for affixes used here could in principle generate a similar dichotomy. If we nominate a decision threshold of, say, 25% for the percentage lowering of fusion rate on affixation (cf. Figure 1), then there appear to be two categories of decomposability: higher decomposability in Group 1 {*-ing* and *-y*} and lower decomposability in Group 2 {*-s* and *-er*}. However, in surveys of other world's languages (e.g., Booij 2005) the dichotomy of affix decomposability has been contested and cases of affixes whose degree of decomposability is stem-dependent have been found.

Rather than testing whether there is such a dichotomy in our results, we chose to investigate the possible explanations for the differences in fusion rates across the four suffixes. These are discussed in Section 3.

3. Variability of McGurk Fusion Rates

The McGurk effect has been studied over the last 3 decades and one question that still remains unanswered is why the fusion rate varies amongst participants and stimuli. Variation is clearly visible in the following studies: in Finnish real and nonce words, fusion rate varied with different stimuli, from 20% to 80% for both nonce words and real words (Sams et al. 1998); in real bisyllabic German words, fusion rates varied from 10% to 100% (Windmann 2004); fusion rates for real English words varied from 29% to 76%, whilst for English nonce words, fusion rate varied from 24% to 74% (Barutchu et al. 2008). In all of these studies, very little explanation is provided for any systematic variation of fusion rate with stimulus. Below, we at least attempt to highlight some strategies that might explain the wide variability in fusion rates across our stimuli.

3.1. *Suffixes beginning with vowel vs. consonant*

In English there are more suffixes that begin with a vowel than with consonants. Most of our stimuli consisted of suffixes with a vowel (*-ing*, *-y*, *-er*) and only one suffix beginning with a consonant (*-s*). It is possible that the fusion rates could be attributable to the type of suffix, V-initial or C-initial. Our results, however, showed no main effect to type of suffix ($\chi^2 = 1.10$, $df = 1$, $p = 0.294$).

3.2. *Vowel context effects and number of syllables*

The possibility that variability of consonant fusion rates in the stem (syllable coda) is attributable to vowel context effects rather than morphological decomposability was examined. Fusion rates were analyzed by stem vowels using classes, open vowels (e.g. ɑ æ ɒ), mid vowels (e.g. ɔ ʌ e) and close vowels (e.g. ɪ i:). A stem that contained open vowels, averaged over all affixes, gave fusion rate of 63.6%, while corresponding averages for mid vowels and close vowels were, respectively 53.8% and close vowels 56.1%. These small differences in average fusion rates are not statistically significant ($\chi^2 = 1.67$, $df = 3$, $p = 0.433$). The results suggest that the gradience which one might attribute to affix is not be related to vowel openness in then stem. This is consistent with earlier studies of the effect of the vowel on McGurk fusion of consonants in CV syllables: contrasting /ɑ/ and /ɪ/ vowels (Green and Gerdeman 1995; Green and Norix 1997; Jordan and Bevan 1997). These earlier studies revealed no significant variation in consonant fusion rates.

We also investigated the possible correlation between the number of syllables in the perceived word (response) and fusion rates. The analysis revealed no correlation between the number of syllables and fusion rates ($r = -0.112$, $p = 0.563$).

4. Discussion

Over the last decade an increasing body of research has started to use the McGurk effect as a tool for probing the mental lexicon. Semantic effects in lexical access and syllable structure have been probed, and now we have begun to probe morphological decomposability. Studies where an incongruent segment is in the coda of syllable CVC or in the syllabic coda of a stem for languages like English (Ali and Ingleby 2004, and the current study; Barutchu et al. 2008), call for care with the perception of coda place of articulation which can sometimes be uncertain.

4.1. *Perceptual cues in onset and coda*

Uncertainties arise because plosives are generally hard to distinguish (especially phoneme /p/ and /t/) on the basis of their burst spectrum. Usually, formant transitions in syllable nuclei are more reliable cues to consonant place than coda spectra (e.g. Liberman et al. 1967, Wright 2001). Therefore it is important to consider whether the alveolar fusion perceived in the syllable coda of the stem by many of the participants in this study is genuine and not simply a confusion of place contrasts (between /p/ and /t/ and between /b/ and /d/) made possible by weak spectral cues in the coda. Our argument against such confusion is based on two points. Firstly, there was no place confusion experienced by participants for congruent stimuli presented randomly amongst our incongruent stimuli. Secondly, our participants were operating audiovisually with visible lip gestures supporting place perception. There is also a body of evidence from other confusability studies. Wright (2001) worked with CV and VC syllables which were presented in clear (no noise) condition. He found little difference in consonant recognition accuracy between onsets and codas. Phoneme /b/ was rarely confused for a /d/ in either the onset or in the coda position. Other studies by Benki (2003) and Benki and Felty (2005) revealed that at various noise levels congruent /p/ rarely elicited a /t/ percept. These studies counter an argument – advanced, e.g., in discussion at the Eleventh Laboratory Phonology Conference, 2008, that the congruently presented phoneme /p/ is often confused with /t/ in presence of noise. To avoid possible place-confusion effects, our stimuli were not presented in noise: in fact, to maximize the signal-to-noise ratio and exclude environmental noise, our audio signal was presented binaurally over headphones rather than through speakers.

Thus, the growing body of evidence on the McGurk effect clearly shows that speech perception is bimodal and humans inherently take into account the visual channel of speech. The audiovisual speech literature supports our contention that the fusion patterns perceived by our participants are not the result of auditory confusion. However, there is a possibility that incongruent stimuli might place a greater cognitive load on the participants and make their perceptions simply erratic. The evidence for cognitive load has been gathered in fMRI studies of brain activation and the degree of integration of audiovisual speech. In one example of such studies, Jones and Callan (2003) found a difference in brain activity brought about by incongruence in audiovisual stimuli. One could probe the cognitive load effect further by comparing decision times for congruent and incongruent audiovisual stimuli. But whether increased cognitive load due to incongruence elicits erratic reporting of percepts is a matter for closer scrutiny. A full investigation of time-limited response to incongruence is needed – with timing (durations) and fusion rates subject to a full multivariate analysis of variance. There is already some evidence that percept reporting is far from erratic, depending systematically on the visual signal.

Clement and Carney (1999) showed that with incongruent audiovisual speech, when the audio signal is degraded, reporting favours the visual signal. This was confirmed with real words with varied noise levels (Ali and Ingleby 2004). We showed that fusion rates for both onsets and codas dropped steadily with noise level and that increased non-fusion responses moved systematically to favour the visual channel rather than distributing erratically over audio and visual channels.

This clearly indicates that in signal-masking conditions, participants rely more on the visual cues. This confirms with the findings of Fixmer and Hawkins (1998), who showed fewer McGurk responses in the presence of visual noise. Unfortunately, their study did not provide data on audio and visual response rates, so the obvious question of whether or not visual signal-masking favours the audio channel remains open. In our experiments reported here, all with clear audio signal, when fusion was not reported, participants favoured the audio channel over the visual channel – a further indication that the acoustic cues in the audio channel are unlikely to have been weak, for if they were weak, we would have seen a shift towards the visual channel such as shown by Clement and Carney (1999).

Little is known about the systematic nature of variation in McGurk fusion rate across stimuli. Currently, we are conducting lip measurements of the visual velar segment (viseme) /k/ and /g/ in our incongruent stimuli to ensure that it is clearly distinguishable from viseme /t/ and /d/ (perceived fusion segment). Previously, with congruent signals, Inverson et al. (1998) have shown in visual-only mode and in audiovisual modality that visual place cues are clearly distinguishable. This suggests that the visual element of the stimulus is clearly distinguishable for place cues. Furthermore, the external lip-shapes corresponding to labial, alveolar and velar mouth gestures have been sketched (Harris and Lindsey 1995) and proffered as the visual cues of underlying place 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**. 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, dichotic fusion and migration, and we shall present more of such frameworks in later papers.

4.2. *Speech perception and morphological decomposability*

It appears from our exploratory study that the quantitative response of fusion rate to morphological context seems a viable alternative option to probing morphology through corpus frequencies. The two methods of probing are more complementary than exclusive. Whilst fusion frequencies are an indicator of the tightness of composition as revealed in earlier studies based on phonetic suprasegmentals such as stress, corpus data reveal the propensity of suffixes to engage in composition. Our results from the fusion rate at stem sites of complex words indicate that *-er, -ing, -s* and *-y* affixes have different degrees of morphological decomposability. Also, our study suggests that the degree of decomposability is gradient rather than dichotomous.

Steady gradience aligns with earlier work by Kuperman et al. (2009) and Hay and Baayen (2005) introducing probabilistic grammar into morphology. These authors argued from corpus evidence that decomposability is inherently graded, as our fusion rates seem to confirm. Others arguing from morphological theory have suggested a categorical distinction. In a larger study with more affixes, it would be possible to correlate the gradience of fusion rate to category membership, using cluster analysis to reconcile the differences between corpus linguists and morphologists. Our present sample of affixes and stems is too small for a definitive reconciliation of opposites, and our coda/onset distinctions have not been fully tested. Currently, we are using the McGurk fusion in a slightly larger study with more morphological affixes and prefixes too. Ideally, we might try dubbing to switch syllable codas and onsets, and begin to evaluate the role played by the stem in decomposability.

The debate regarding affix ordering and decomposability continues, for example, at the workshop on Affix Ordering in Typologically Different Languages during the 13th International Morphology Meeting and at the Berkeley Workshop on Affix Ordering, both in 2008. The work of Hay and Plag (2004) uses frequency-based evidence and focuses only on English suffixes, while in Kuperman et al.'s (2009) study corpus frequencies and eye-tracking data are employed. A large-scale clustering study across more- and less-agglomerative languages would settle some issues, since the degree of decomposability in morphology is possibly an adjunct of greater exposure of participants to polymorphic combinations. A form of evidence that we have not discussed here, but should probably figure in any large-scale MANOVA of affixing is perception errors. Studies of these have tended to align decomposability with grammarian's categories: for example Janssen and Humphreys (2002) note that inflectional suffixes are more susceptible to perceptual errors than derivational suffixes.

In this paper, we have used an experimental approach with incongruent audiovisual speech data. The longer term aim is to extend this research to other languages, on a parallel to our earlier studies probing the very different cases of syllabic structure in English and Arabic. Arabic, a Semitic language, Dutch, a Germanic language, and a Slavic language would be useful to free any provisional results of the scourge of monolingualism. For example, incongruent segments in the syllable coda of stems for different morphological affixes are shown below in (26) for Arabic and in (27) and (28) for Dutch.

(26) $_{A}$(sa**bb**áhumaa || sa**kk**áhumaa)$_{V}$ → (sa**dd**áhumaa)$_{F}$, ...
$_{A}$(سَبَّهُمَا || سَگَّهُمَا)$_{V}$ → (سَدَّ هُمَا)$_{F}$, ...

(27) $_{A}$(rʊ:**p**ən || rʊ:**k**ən)$_{V}$ → (rʊ:**t**ən)$_{F}$, ...

(28) $_{A}$(slɔɪ**p**ər || slɔɪ**k**ər)$_{V}$ → (slɔɪ**t**ər)$_{F}$, ...

5. Conclusion

This first exploration of morphological decomposability using McGurk fusion to probe segments in the syllable coda of stems has exposed some systematic patterns, and is definitely worth continuing further with a large corpus of items. We have been able to rank the degree of decomposability of compounds in a few common English suffixes to stems, using the lowering of fusion rate as an indicator. Although we have used audiovisual incongruent stimuli, the approach could be made entirely acoustic by using dichotic incongruence and stereophonic techniques, as others have done in migration studies. Methods using incongruent data and fusion responses contrast with those based on corpus frequencies and on the effect on suprasegmentals such as stress.

The morpheme boundaries that we have examined have been at the syllable coda of stems, and there has been concern in the past about the reliability of place of articulation perceptions in codas. Acoustic experiments on /p/ and /t/, /b/ and /d/ confusion in codas show high confusability in the presence of noise, but we have avoided such problems by using high quality audio and low-noise headphones, and using a visual channel showing lip gestures to reinforce place perception.

There is a need to go beyond the current exploratory study and look at larger numbers of suffixes and a variety of stems. The ranking of decomposability in English is likely to differ significantly from that in other languages where templatic and agglomerative morphology carries a greater proportion of the information in compound words. We have indicated some desirable features of such a large-scale study in Section 4.

Appendix 1 – Incongruent word stimuli tested and response rate

Audio	Visual	Expected & Perceived Fusion	Audio Response %	Visual Response %	Fusion Responses %	Non-word response %*
/bɪbɪŋ/	/bɪgɪŋ/	/bɪdɪŋ/	0	0	100	
/tæbɪŋ/	/tægɪŋ/	/tædɪŋ/	12	0	88	
/nɒbɪŋ/	/nɒgɪŋ/	/nɒdɪŋ/	0	12	88	
/tʃiːpɪŋ/	/ tʃiːkɪŋ/	/tʃiːtɪŋ/	0	25	75	
/rɪbɪŋ/	/rɪgɪŋ/	/rɪdɪŋ/	13	12	75	
/reɪpɪŋ/	/reɪkɪŋ/	/reɪtɪŋ/	0	28	72	
/wɪpɪŋ/	/wɪkɪŋ/	/wɪtɪŋ/	14	14	72	
/klɒpɪŋ/	/klɒkɪŋ/	/klɒtɪŋ/	0	37	63	
/sliːpɪŋ/	/sliːkɪŋ/	/sliːtɪŋ/	43	0	43	14
/kəʊpɪŋ/	/kəʊkɪŋ/	/kəʊtɪŋ/	94	6	0	
/wɪpə/	/wɪkə/	/wɪtə/	12	0	88	
/flɪpə/	/flɪkə/	/flɪtə/	25	12	63	
/rʌbə/	/rʌgə/	/rʌdə/	12	50	38	
/slɪpə/	/slɪkə/	/slɪtə/	69	19	12	
/bɪbə/	/bɪgə/	/bɪdə/	50	38	12	
/kæbɪ/	/kækɪ/	/ kædɪ /	0	0	100	
/bæbɪ/	/bægɪ/	/bædɪ/	13	12	75	

Audio	Visual	Expected & Perceived Fusion	Audio Response %	Visual Response %	Fusion Responses %	Non-word response %*
/dɪpɪ/	/dɪkɪ/	/dɪtɪ/	15	14	57	14
/hɒpɪ/	/hɒkɪ/	/hɒtɪ/	50	14	36	
/bɒbɪ/	/bɒgɪ/	/bɒdɪ/	50	13	12	25
/tɪps/	/tɪks/	/tɪts/	27	27	46	
/tæbz/	/tægz/	/tædz/	14	14	72	
/kɒbz/	/kɒgz/	/kɒdz/	27	9	64	
/mæps/	/mæks/	/mæts/	64	9	27	
/peps/	/peks/	/pets/	41	0	59	
/kɔːps/	/kɔːks/	/kɔːts/	36	0	64	
/pʌps/	/pʌks/	/pʌts/	0	27	73	
/bɪbz/	/bɪgz/	/bɪdz/	36	9	55	
/bʌbz/	/bʌgz/	/bʌdz/	37	36	27	

* Responses reported by participants were nonwords, e.g. /sliːθɪŋ/, /dɪθɪ/ and /bɒðɪ/.

Correspondence e-mail address: dr.aza.ali@gmail.com

Notes

1. Giegerich (1999) distinguishes English affixes as Class 1 and Class 2, although the classification is debatable.
2. For future study, we aim to design a set of stimuli (stem + affix) that are nonce words, to determine further whether fusion rates are the same across different suffixes.
3. Double-blinding uses an administrator to label the stimuli so that neither the participants nor the experimenter know which stimuli are congruent or incongruent audiovisual stimuli. The labeling code is revealed only after compilation of results.
4. The reduction is tied to morphological complexity not to the changed position of incongruent segment brought about by affixing: in our previous studies, effect of change of position of incongruent segment in branching onsets and in branching codas was found to have no significant impact on fusion rate.

References

Ali, Azra, Ashraf Hassan-Haj, Michael Ingleby & Ali Idrissi. 2005. McGurk fusion effects in Arabic words. In Eric Vatikiotis-Bateson (ed.), *Auditory-Visual Speech Processing 2005*, 11–16. British Columbia, Canada (http://www.isca-speech.org/archive/avsp05).

Ali, Azra & Michael Ingleby. 2004. Probing cognitive mental models of speech using the McGurk effect. Paper presented at *Fourth international conference on the mental lexicon.* Windsor, Ontario, Canada.

Ali, Azra, Michael Ingleby & David Peebles. 2008. Anglophone perceptions of Arabic syllable structure. Paper presented at *CUNY conference on the syllable*. New York, USA.

Baothman, Fatmah & Michael Ingleby. 2002. Representing coarticulation processes in Arabic speech. In Sami Boudelaa (ed.), *Perspectives on Arabic Linguistics XVI*, 95–102. Amsterdam: John Benjamin.

Barutchu, Ayla, Sheila Crewther, Patricia Kiely, Melanie Murphy & David 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). 1–11.

Benkí, Jose. 2003. Analysis of English nonsense syllable recognition in noise. *Phonetica* 60. 129–157

Benkí, Jose & Robert Felty. 2005. Recognition of English phonemes in noise. *Journal of the Acoustical Society of America* 117. 2568.

Booij, Geert. 2005. The interface between phonology and morphology. A comment on Ellen Kaisse, 'Word formation and phonology', *Skase Journal of Theoretical Linguistics* 2. 17–25.

Brancazio, Lawrence. 2004. Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 30(3). 445–463.

Clement, Bart & Arlene Carney. 1999. Audibility and visual biasing in speech perception. *Journal of Acoustical Society of America* 106(4:2). 2271.

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

Cutler, Anne, Jacques Mehler, Dennis Norris & Juan Segui. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25. 385–400.

Cutting, James. 1975. Aspects of phonological fusion. *Journal of Experimental Psychology: Human Perception and Performance* 104(2). 105–120.

Cutting, James. 1976. Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review* 83. 114–140.

Dixon, Robert. 1977. *A grammar of Yidin*. Cambridge: Cambridge University Press.

Fixmer, Eric & Sarah Hawkins. 1998. The influence of quality of information on the McGurk effect. In Denis Burnham, Jordi Robert-Ribes & Eric Vatikiotis-Bateson (eds.), *Auditory-Visual Speech Processing (AVSP '98),* 27–32. (http://www.isca-speech.org/archive/avsp98).

Gelder, Beatrice de, Paul Bertelson, Jean Vroomen & Hsuan Chen. 1995. Interlanguage differences in the McGurk effect for Dutch and Cantonese listeners. *Fourth European Conference on Speech Communication and Technology.* 1699–1702.

Giegerich, Heinz. 1999. *Lexical strata in English*. Cambridge: Cambridge University Press.

Green, Kerry & Anthony Gerdeman. 1995. Cross-modal discrepancies in coarticulation and the integration of speech information: The McGurk effect with mismatched vowels. *Journal of Experimental Psychology: Human Perception and Performance* 21(6). 1409–1426.

Green, Kerry & Linda Norrix. 1997. Acoustic cues to place of articulation and the McGurk effect: The role of release bursts, aspiration and formant transitions. *Journal of Speech, Language, and Hearing Research* 40. 646–665.

Halwes, T. G. 1969. *Effects of dichotic fusion on the perception of speech*. Unpublished PhD thesis, University of Minnesota.

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

Hay, Jennifer. 2003. *Causes and Consequences of word structures*. London & New York: Routledge.

Hay, Jennifer & Harald Baayen. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* 9. 342–348.

Hay, Jennifer & Ingo Plag. 2004. What causes possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory* 22(3). 565–596.

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

Janssen, Dirk & Karin Humphreys. 2002. Morphological speech errors on agentive and comparative affixes. *Poster Presented at the Third International Conference on the Mental Lexicon*. Max Bell Building, Banff, Canada.

Jones, Jeffery A. & Daniel Callan. 2003. Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport* 14(8). 1129–1133.

Jordan, Timothy & Kim Bevan. 1997. Seeing and hearing rotated faces: Influences of facial orientation on visual and audiovisual speech recognition. *Journal of Experimental Psychology: Human Perception and Performance* 23(2). 388–403.

Kaisse, Ellen. 2005. Word formation and phonology. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of English Word Formation,* 25–47. Dordrecht: Springer.

Kolinsky, Regine, Jose Morais & Mireille Cluytens. 1995. Inter-mediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language* 34. 19–40.

Kuperman, Victor, Robert Schreuder, Raymond Bertram & Harald Baayen. 2009. Reading Polymorphemic Dutch Compounds: Towards a Multiple Route Model of Lexical Processing. *Journal of Experimental Psychology: Human Perception and Performance* 35(3). 876–895.

Liberman, A. M., F. S. Cooper, D. P. Shankweiler & M. Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review* 74(6). 431–461.

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

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

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

Sams, Mikko, Petri Manninen, Veikko Surakka, Pia Helin & 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 & Ludovic Ferrand. 2002. The Role of the Syllable in Speech Perception and Production. In Jacques Durand & Bernard Laks (eds.), *Phonetics, Phonology and Cognition,* 151–167. Oxford: Oxford University Press.

Shaker, Anshula, Brooke Hallowell, Robert Roe, Hans Kruse, Danny Moates & Chao-Yang Lee. 2006. Using Eye Movement Indices to Capture Semantic Priming Effects. http://convention.asha.org/2006/handouts/855_1791Odekar_Anshula_056074_102906044854.ppt. (4 April 2008).

Taft, Marcus. 1979. Lexical access via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning & Verbal Behavior* 18. 21–39.

Tiippana, Kaisa, Mikko Sams & Riikka Möttönen. 2000. The effect of sound intensity and noise level on audiovisual speech perception. *Poster Presented at the Second International Multisensory Research Forum* 36. Tarrytown, New York, United States of America. http://imrf.mcmaster.ca/IMRF/2000/abstracts.html#a4

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 Elizabeth Hume & Keith Johnson (eds.), *The role of the interplay of speech perception and phonology,* 251–277. New York: Academic Press.