Psycholinguists Should Resist the Allure of Linguistic Units As Perceptual Units

Arthur G. Samuel [a,b,c]

[a] *Basque Center on Cognition, Brain and Language, Donostia, Spain.*

[b] *IKERBASQUE, Basque Foundation for Science.*

[c] *Stony Brook University, Dept. of Psychology, Stony Brook, NY, the United States of America.*

Corresponding author:

Arthur G. Samuel

Department of Psychology

Stony Brook University

Stony Brook, NY 11794-2500

a.samuel@bcbl.eu

Keywords: perceptual units; linguistic units; selective adaptation in speech

ABSTRACT

The current study has empirical, methodological, and theoretical components. It draws heavily on two recent papers: Bowers et al. (2016) (*JML*, *87*, 71-83) used results from selective adaptation experiments to argue that phonemes play a critical role in speech perception. Mitterer et al. (2018) (*JML*, *98*, 77-92) responded with their own adaptation experiments to advocate instead for allophones. These studies are part of a renewed use of the selective adaptation paradigm. *Empirically*, the current study reports results that demonstrate that the Bowers et al. findings were artifactual. *Methodologically*, the renewed use of adaptation in the field is a positive development, but many recent studies suffer from a lack of knowledge of prior adaptation findings. As the use of selective adaptation grows, it will be important to draw on the considerable existing knowledge base (this literature is also relevant to the currently popular research on phonetic recalibration). *Theoretically*, for a half century there has been a recurring effort to demonstrate the psychological reality of various linguistic units, such as the phoneme or the allophone. The evidence is that listeners will use essentially any pattern that has been experienced often enough, not just the units that are well-suited to linguistic descriptions of language. Thus, rather than trying to identify any special perceptual status for linguistic units, psycholinguists should focus their efforts on more productive issues.

Psycholinguists Should Resist the Allure of Linguistic Units As Perceptual Units

In his thoughtful and intriguing article, Elman (2009) began by saying: “I begin with a warning to the reader. I propose to do away with one of the objects most cherished by language researchers: the mental lexicon. I do not call into question the existence of words, nor the many things language users know about them. Rather, I suggest the possibility of lexical knowledge without a lexicon.” (Elman, 2009, p. 548). Following this precedent, I also begin with a warning: I propose to do away with a cherished endeavor of psycholinguists. Despite the clear utility of linguistic units in describing language (the core purpose of linguistic analysis), attempts by psycholinguists to demonstrate “the psychological reality of X”, where “X” is some linguistically well-motivated unit, have repeatedly been fruitless. It is not that linguistic units cannot be used by listeners; rather, it is that almost any often-encountered pattern can be, whether that pattern corresponds to a linguistic unit or not. As such, demonstrating that (some) linguistic units can (sometimes) be used by listeners does not significantly advance our understanding of speech perception.

My call to abandon the search for linguistically-defined perceptual units is not new. Almost 20 years ago, Goldinger and Azuma (2003), in a paper entitled “Puzzle-solving science: The quixotic quest for units in speech perception”, made essentially the same point. Moreover, they noted that 30 years before their own paper, researchers (e.g., Foss & Swinney, 1973; McNeill & Lindig, 1973; Savin & Bever, 1970) had already

begun to raise related concerns. As Goldinger and Azuma put it, “Considered collectively, 30 years of speech-unit research has generated little apparent progress. If the goal was to decide a ‘‘winner’’, the enterprise has clearly failed: Despite dozens of studies, the candidate list has actually grown… [T]he classic question of speech units seems ill conceived.” (p. 307).

Despite this insightful analysis, the effort to reify linguistic units is alive and well. In fact, the current study was stimulated by two recent papers in which the goal was to provide evidence that perceptual processing of language relies on particular linguistic units. The first paper, by Bowers, Kazanina, and Andermane (2016) (hereafter, BKA16), made an emphatic claim for the phoneme as an important perceptual unit during spoken word recognition (a claim that was then even more strongly asserted in a follow-up paper by Kazanina, Bowers, & Idsardi, 2018). The second paper, by Mitterer, Reinisch, and McQueen (2018) (hereafter, MRM18), argued forcefully for the allophone’s primacy over the phoneme as a perceptual unit, in a rebuttal to BKA16. These papers are clear examples of the approach that I am highlighting, as the core research goal in each paper was to make a very strong argument in favor of a particular linguistic unit – the phoneme for BKA16, and the allophone for MRM18. As I will expand on below, this type of research goal has consistently proven to be fruitless.

In addition to making this theoretical point, there are two other goals of the current study. The first is empirical: The experiments in the current study test whether the key result reported by BKA16 is an artifact of the stimuli that they used. If the result is artifactual, then of course the conclusions drawn from it are baseless. The other goal of the current study is more methodological: As I will discuss shortly, the technique used

by BKA16 and by MRM18 is a variation of a methodology that was widely used 40 years ago, and that is enjoying renewed popularity. One goal of the current study is to urge new users of the technique to thoroughly acquaint themselves with the methodological and empirical findings in the original literature. As noted below, familiarity with this literature is also important for researchers examining phonetic recalibration (also called “retuning” or “perceptual learning”) because many of the studies in this burgeoning area have substantial overlap methodologically with selective adaptation procedures.

The research by BKA16 and MRM18 was conducted using a modified version of the selective adaptation procedure. In order to fully understand their work, some knowledge of that task is necessary because the empirical questions examined in those papers were based on earlier adaptation results. Therefore, I will present a very brief overview of the adaptation literature. A more extensive description of much of the relevant work is available in Samuel (1986).

The seminal selective adaptation paper was done by Eimas and Corbit (1973). Research on visual psychophysics had shown that repeatedly experiencing a stimulus reduced sensitivity to a visual property. Eimas and Corbit took this approach into the field of speech perception by first creating a continuum of syllables that ranged from /ba/ to /pa/, and then testing whether perception of those syllables would be changed by repeated exposure to an endpoint sound. They observed a “selective adaptation” effect akin to what had been found in visual psychophysics: They reported that after hearing /ba/ many times, listeners became less sensitive to /ba/ and heard fewer /ba/ sounds than before exposure; after hearing /pa/ many times, listeners became less sensitive to /pa/ and heard fewer /pa/ sounds than before. Borrowing from theories in visual

psychophysics, Eimas and Corbit interpreted the effect as being a consequence of fatiguing “linguistic feature detectors” through repeated stimulation.

The Eimas and Corbit (1973) findings, together with other similar papers that appeared soon after, generated enormous interest and research in the field of speech perception. Adaptation rapidly became a favorite technique, and it was used to look at a range of questions. However, within a few years, a number of authors questioned the idea that there were linguistic feature detectors being fatigued (e.g., Diehl, 1981; Diehl, Elman, & McCusker, 1978). The general thrust of the alternative was that the observed shifts were not based on fatigue, but were instead a manifestation of a more general contrast effect that had to do with decision-making rather than perception. Even though one could make the argument that this issue does not actually undercut the utility of the paradigm (see Samuel, 1986 for such an argument), by the late 1980’s the technique had fallen out of favor.

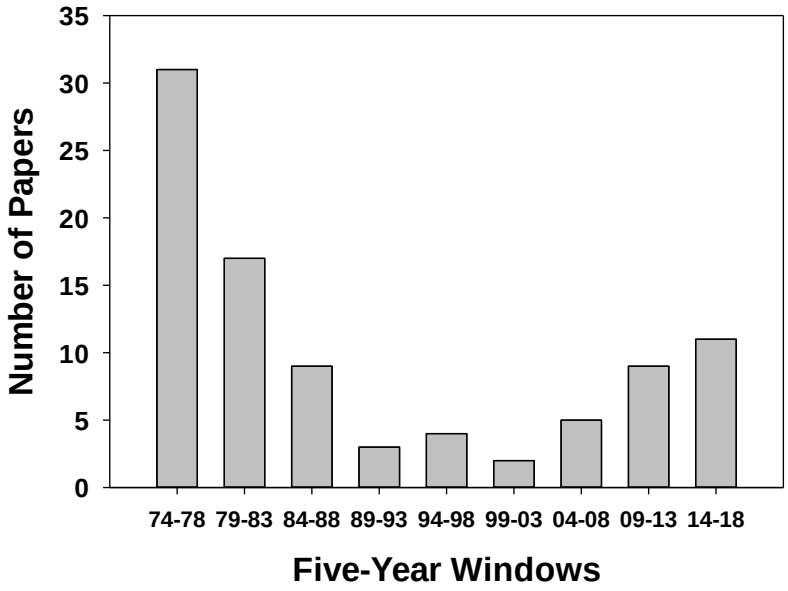

Figure 1*: Number of empirical selective adaptation papers published in five-year windows, following the seminal 1973 paper by Eimas and Corbit. There was an initial boom, followed by a 20-year drought, with a new growth in the use of the technique in the last 10-15 years.*

Figure 1 illustrates the boom-bust-boom pattern of speech research using the selective adaptation paradigm. The numbers here are approximations, based on the number of empirical papers in a given year that cited the original Eimas and Corbit (1973) paper. As such, they slightly underestimate the number of papers because not all papers cite the seminal work, an undercount that presumably increases in more recent years. The pattern is very clear: After a big boom during the five years following Eimas and Corbit, the task started to be used less and less in each of the successive five-year windows. Starting in the late 1980's, for about 20 years, there was very little use of the technique.

During the last 10-15 years there has been a clear rebound, with growing use of adaptation again. What underlies this resurgence? It appears that much of adaptation's renewed appeal stems from its potential relationship to a different phenomenon that has generated great interest in the field – perceptual recalibration. The two seminal papers on perceptual recalibration were by Norris, McQueen, and Cutler (2003), and by Bertelson, Vroomen, and de Gelder (2003), both appearing just before the rebound in selective adaptation research. Norris et al. showed that when an

ambiguous segment (e.g., midway between /s/ and /f/) is presented in a number of lexical contexts that disambiguate it, listeners show evidence of expanding the phonemic category to include the ambiguous sound. Bertelson et al. showed a corresponding phoneme category expansion when the disambiguation comes from lipread information. Although the recalibration effect goes in the opposite direction than adaptation (i.e., boundaries shift to expand a phonemic category, versus a decrease with adaptation), both involve a shift in category boundaries through exposure to speech input. In fact, in many of the recalibration experiments that Vroomen and his colleagues have run (e.g., Vroomen, van Linden, de Gelder, & Bertelson, 2007; Vroomen, van Linden, Keetels, de Gelder, & Bertelson, 2004), there is an explicit comparison of adaptation and recalibration. The apparent similarities have been captured in a Bayesian model that uses the same formal procedures to model both effects (Kleinschmidt & Jaeger, 2015).

Although adaptation can be an extremely useful tool (see Samuel, 1986), many of the new selective adaptation and recalibration papers are not well-informed about much of the work done during the original “boom”. There are unmotivated and unexplained changes in procedures, and in many cases, a lack of knowledge about what has already been established. There are about a hundred empirical adaptation papers in the literature (see Figure 1), in many cases with multiple experiments, meaning that there are hundreds of prior adaptation experiments. New studies using adaptation rarely are well-informed by this literature, and this problem is even worse in recalibration studies that use procedures that effectively and often unknowingly create adaptation situations. Sticking to adaptation itself, the two papers that are the focus

here are typical: BKA16 cited only three adaptation papers among their approximately 50 citations. MRM18, with nearly 80 citations, mentioned five adaptation papers; of these, one was the seminal Eimas and Corbit (1973) paper, and two of the other four were actually papers that question the utility of adaptation. As I will discuss in the General Discussion, there are many results in the adaptation literature that are relevant to the research in these two papers.

I will initially focus on the BKA16 paper because the empirical part of the current study is directly based on that study. Recall that the core theoretical claim by BKA16 was that the phoneme is a critical perceptual unit for listeners. They framed their effort as follows: "Traditional linguistic theory postulates a small set of phonemes that can be sequenced in various ways in order to represent thousands of words in a language …The common rejection of position invariant phonemes in psychological theories and models of word perception is a fundamental claim, and we explore this issue here… [W]e describe two experiments that provide strong evidence that phonemes do indeed play a role in word perception." (pp. 71-72).

For the two experiments in their study, the key phrase in their framing was "position invariant phonemes". Phonemes are the vowels and consonants of a language, and in linguistic analyses, their positional invariance is an important property. This property can be illustrated with the phoneme /p/ in English. Although most speakers are not aware of it, there is systematic variation in the way that /p/ is produced in English words. Specifically, when the /p/ is an onset (e.g., in "park"), the /p/ is aspirated – there is a puff of air after the lips open; in contrast, when the /p/ is part of a cluster (e.g., in "spark"), the /p/ is unaspirated – there is no puff of air. The two variants

(aspirated and unaspirated /p/) are called allophones because they both are members of the broader /p/ phonemic category. Essentially, phonemes are abstractions across the relevant allophones. Because BKA16 were arguing that phonemes are critical perceptual units, perceptual tests should show that listeners treat phonemes as the same, despite differences in position.

From this perspective, one of the early papers in the adaptation literature posed a significant problem: Ades (1974) demonstrated that syllable-initial phonemes produced adaptation on syllable-initial test items (e.g., /bæ/ shifted identification of members of a /bæ/-/dæ/ continuum), and syllable-final phonemes produced adaptation on syllable-final test items (e.g., /æb/ shifted identification of members of an /æb/-/æd/ continuum), but there was no adaptation when adaptors and test syllables differed in position (e.g., /bae/ did not affect identification of /æb/-/æd/ test items). Samuel (1989) reported the same positional specificity for adaptation. BKA16 mentioned these two papers in their review of research that they saw as potentially problematic for their argument that phonemes are essential perceptual units, and the empirical portion of their paper was designed to demonstrate that adaptation actually does occur despite positional mismatching of the adaptors and test items.

Although there are of course variations across the many studies in the adaptation literature, there are certain procedures that are most common. In a typical study, simple consonant-vowel or vowel-consonant stimuli serve as both the adaptors and the test items. Often there will be 6-8 test items that form a continuum (e.g., with a good “ba” at one end, and a good “da” at the other), and the adaptors are usually the continuum endpoints or stimuli chosen to have a particular relationship to them. For example, with

a /ba/-/da/ test series, in addition to the endpoint /ba/ and /da/ sounds, adaptors could be /pa/ and /ta/, chosen to share the place of articulation difference of the endpoints, but to differ from them in voicing. Typically, an adaptation study includes 10-20 cycles, with each cycle including about 30-60 seconds of hearing a repeating adaptor followed by listeners identifying one randomization of the test continuum items. This procedure produces a psychometric function (the probability of identifying each continuum item as one of the two categories) for each adaptation condition. Often there is also a psychometric baseline function, based on identifying the test items before any adaptation occurs. Adaptation manifests as a shift of one psychometric function relative to another, usually with the largest shift near the middle of the continuum.

The task used by BKA16 maintained the core properties of repeating a sound (the adapting sequence) and identification of test items that are midway between two good endpoints. However, the implementation was quite different. For simplicity, I will focus on their test involving a contrast between /b/ and /d/; there was also a test involving /s/ and /f/, but that part of their study is not particularly relevant to the core issues, or to the experiments in the current study. Rather than having listeners identify items that spanned a continuum, BKA16 had them identify a single token that was selected to be midway between “bump” and “dump”. Instead of repeating an endpoint item as the adaptor, BKA16 played listeners sets of words that had the adapting sound in a particular position. For example, there were 25 words that all started with /b/ (e.g., “bail”, “bank”, “berry”, “bother”…), and 25 words that all started with /d/ (e.g., “dice”, “draft”, “donkey”, “driver”…). For a within-position adaptation test, these word-initial items were presented (a number of times), and listeners identified the ambiguous

“bump-dump” stimulus a number of times. For a between-position test, 25-item word sets with final-position critical sounds (e.g., “curb”, “glib”, “cherub”, “reverb” for /b/, and “gold”, “need”, “lucid”, “salad” for /d/) were used as the repeating items (there were also medial-position items, which are not relevant to the current study). All stimuli were based on recordings made by a native speaker of British English.

These procedures did produce significant adaptation effects, measured by differences in how people identified the ambiguous “bump-dump” token as a function of whether the repeated words included /b/ or /d/. Critically for BKA16’s argument, there was a significant shift for the final-position adapting words on the initial-position test item. The cross-position effect was smaller than the matched-position effect (about one third as large looking at all subjects, or about one half as large for subjects without ceiling/floor effects), but it was significant. BKA16 took this cross-position adaptation as evidence for position-invariant phonemes.

Given this claim, they said “[W]e would note that there is one set of findings that does seem at odds with our results; namely, the previous adaptation studies that failed to obtain adaptation across syllable positions in non-lexical targets (Ades, 1974; Samuel, 1989). Why the difference?” (p. 79). They answered this question by offering three “speculative explanation[s]”. One suggestion was that using a single ambiguous test token might be more sensitive than using a full continuum of test sounds. This does not make sense, as a full continuum includes an ambiguous region. Moreover, BKA16 had to drop over half of their subjects to get their cleaner measure, precisely because subjects differ in exactly where the ambiguous region will be (by testing the full range, this problem is mitigated). The individual differences in phonetic boundaries are

usually within the ambiguous region of a continuum, allowing most listeners to be included in the data analyses. Their second speculation was that their lexical adaptors might be producing additional adaptation at a lexical level. However, Samuel (1997, experiments 3A and 3B) demonstrated that there is no contribution to adaptation at the lexical level itself. Finally, BKA16 said "In addition, whether or not our procedure is better suited for accessing abstract phoneme representations, the important point to emphasize is that the previous authors relied on null results in their adaptation studies to reject phonemes." (p. 80).

In general, of course, caution is called for in accepting a null effect. However, null effects can indeed exist, and when multiple tests yield null effects, at some point accepting the null is the correct decision. If in fact there were only two tests that yielded null effects, accepting the null might well be premature. However, the evidence against cross-position adaptation is much more substantial. In addition to the Ades (1974) and Samuel (1989) papers cited by BKA16, the positional-specificity issue was tested by Sawusch (1977b), Wolf (1978), and Samuel, Kat, and Tartter (1984). Table 1 summarizes the results from the five studies. Across these studies, there were 18 within-position tests, and all 18 produced significant adaptation effects. There were 18 across-position tests, and 14 of these failed to find adaptation (for 14 out of 18, $p < .05$ by a sign test). Thus, there is extremely substantial evidence for positional specificity for adaptation.

The four significant cases of across-position adaptation in the prior literature are themselves informative. Two of these came from Wolf's study, and Wolf included identical noise bursts across initial and final position stimuli. The other two came from

Samuel's (1989) test of liquids; to make convincing liquids Samuel included identical 70 msec steady state formants in initial and final position. Thus, for the few cases of (weak) cross-position adaptation (versus the large majority of null effects), the adaptation was almost certainly due to the stimuli including strong acoustic matches across position, rather than to the positions sharing phonemic identity.

POSITION – MATCHED			**POSITION - MISMATCHED**		
Test Series	**Adaptors**	**Effect**	**Test Series**	**Adaptors**	**Effect**
		Ades (1974)			
bæ-dæ	bæ	**Signif**			
	dæ	**Signif**			
æb- æd	æb	**Signif**			
	æd	**Signif**			
			bæ - dæ	æb	**No effect**
				æd	**No Effect**
			æb - æd	bæ	**No effect**
				dæ	**No effect**
		Sawusch (1977b)*			
bæ - dæ	bæ	**Signif**			
ʌb - ʌd	ʌd	**Signif**			
			bæ - dæ	ʌd	**No effect**
			ʌb - ʌd	bæ	**No effect**
		Wolf (1978)**			
dæ - gæ	dæ	**Signif**			
dæ - gæ	gæ	**Signif**			
æd - æg	æd	**Signif**			
æd - æg	æg	**Signif**			
			dæ - gæ	æd	Small

			dæ - gæ	æg	**No Effect**
			æd - æg	dæ	**Signif**
			æd - æg	gæ	**No Effect**

Samuel, Kat, & Tartter (1984)

ba-da	ba	**Signif**			
ab-ad	ad	**Signif**			
			ba-da	ab	**No Effect**
			ab-ad	ba	**No Effect**

Samuel (1989)***

ba-da	ba	**Signif**			
ab-ad	ad	**Signif**			
			ba-da	ab	**No Effect**
			ab-ad	ba	**No Effect**
ri-li	ri	**Signif**			
ri-li	li	**Signif**			
ir-il	ir	**Signif**			
ir-il	il	**Signif**			
			ri-li	ir	**No Effect**
			ri-li	il	**No Effect**
			ir-il	ri	Small
			ir-il	li	Small

Effects labeled “Small” were significant but also much smaller than within-position shifts.

*Sawusch (1977b) used different vowels in order to be able to have formant transitions, across positions, that were acoustically identical.

**Wolf (1978) included noise bursts that were identical for the onset of CV and offset of VC stimuli.

***Samuel (1989) included 70 msec steady-state formants that were identical for the onset of CV and offset of VC stimuli.

***Table 1**: Pattern of shifts and failures to shift as a function of whether adaptors and test sounds matched in position.*

The handful of small but significant acoustically-driven cross-position adaptation effects raises the question of whether there might be a similar source for the small but significant effects reported by BKA16. Recall that the adaptor words were items that were recorded by a native speaker of British English. In British English, especially for the citation-form speech recorded for research, a native speaker is likely to produce a "released" final stop consonant. In a released final stop, rather than simply end the word with the stop closure, the speaker releases the closure to produce a more clearly articulated sound. Critically, such a release is acoustically largely the same as the normal articulation of that stop consonant in initial position. If the final-position adapting words had many released stops then listeners would be receiving acoustic input that matches the onset of the "bump-dump" test item. In fact, in Footnote 3 (p. 75), BKA16 report that 11 of the 25 /b/ adaptors had released final stops, and all 25 of the final /d/ adaptors did.

The presence of released final stops in most of the adaptors prompts an obvious question: Were the observed shifts due to the resulting acoustic matching across position, rather than to shared phonemic representations? There is a straightforward way to answer this question: Adaptation can be conducted using the original (released) adaptors, and with versions of those adaptors in which the releases have been spliced off the ends of the words. Those two tests are reported in Experiment 1; Experiment 2 reports the results of two control conditions.

EXPERIMENT 1

As noted above, the procedures used by BKA16 differed in several ways from what is typically done in selective adaptation experiments. In the current study, we use procedures that are more in line with standard practice. The most important change is that rather than having subjects repeatedly judge the identity of a single token (a token taken from a continuum between “bump” and “dump”), listeners in the current study identified members of an 8-step continuum ranging between /ba/ and /da/. Following previous practice (e.g., Samuel, 1989, 2016), statistical analyses are based on the identification of the middle four members of the continuum. This approach focuses on the ambiguous region (as BKA16’s single token is intended to do), while still being sensitive to effects despite individual differences in listeners’ phoneme boundaries (recall that to remove floor and ceiling effects, BKA16 needed to discard over half of their subjects due to individual differences in boundary location).

Using these procedures, two conditions were tested in Experiment 1. In one condition (Original), the original adapting words with word-final /b/ or /d/ used by BKA16 served as the adaptors. In the second condition (No-Release), edited versions of these stimuli were used as the adaptors. For this condition, the ending of each original adaptor word was carefully examined (and listened to) using a waveform editor, and words were cut just before any release. Given these two sets of stimuli, Experiment 1 tests two questions: (1) Will the original stimuli produce significant cross-position

adaptation shifts on the /ba/-/da/ test items? (2) If so, will those shifts still be found for the stimuli that do not have any release-based acoustic cues?

Method

Participants

A total of 53 participants took part in Experiment 1, 27 with the Original stimuli, and 26 with the No-Release adaptors. All were native speakers of American English, with no self-identified hearing problems. They received credit toward a course requirement for their participation.

Stimuli

Test syllables: An 8-step consonant-vowel (CV) /ba/-/da/ test continuum was used. The stimuli came from the same 10-step test series that provided the stimuli used by Samuel (1989); the subset of 8 items was shifted (i.e., items 2 through 9, rather than 1 through 8) to better center the continuum. All stimuli were 240 msec long, of which the first 40 msec consisted of formant transitions and the final 200 msec were steady-state vowel transitions. The items had been generated using the cascade branch of the Klatt synthesizer, using a fundamental frequency range (between 100 and 140 Hz) typical of a male voice. See Samuel (1989) for more details of the synthesis.

Adaptors: In the Original condition, the adaptors were the 25 final-/b/ and the 25 final-/d/ words used by BKA16. These had been recorded by a male speaker of British English, Received Pronunciation. In the No-Release condition, the same 50 words were used as adaptors, but each had been trimmed to remove any release that had

been present. Table 2 lists the adapting words, and shows both the original and trimmed durations for each.

/b/ Adaptors	Original Duration	Trimmed Duration	/d/ Adaptors	Original Duration	Trimmed Duration
arab	415	285	ahead	526	443
cherub	497	391	avid	442	365
club	445	313	cord	495	417
crib	461	345	gold	541	464
curb	529	426	humid	612	531
glib	427	264	liquid	595	508
globe	551	481	load	582	504
grab	458	384	lucid	598	516
grub	487	397	need	639	557
herb	556	489	orchard	567	449
lobe	580	514	plod	422	369
perturb	700	586	pond	611	536
probe	515	457	reed	631	558
proverb	652	591	road	627	554
reverb	674	571	salad	558	500
robe	575	516	spade	579	524
scrub	592	498	tend	587	502
shrub	529	446	timid	401	306
slob	583	461	tread	509	378
snob	587	465	vivid	600	486
superb	723	596	weed	528	466
throb	597	454	wicked	459	383
tribe	633	557	wind	655	576
tube	591	501	word	616	541
verb	674	565	yard	599	524
MEAN	**561**	**462**	**MEAN**	**559**	**478**

Table 2*: Adaptor words with final /b/ or with final /d/. The original duration of each word, and the trimmed duration of each word, is specified in milliseconds. On average, words were trimmed by 90 msec.*

Apparatus and Procedure

Participants were tested in sound-shielded chambers, with 1-3 participants tested at a time. They listened to the speech over SONY MDR-V900 stereo headphones, and responded by using two labeled buttons on a response pad in front of them.

Participants came to the lab twice, with the second session run two days after the first session. Each session included an initial identification task, followed by an adaptation task. The adaptors during one session were the final-/b/ words, and during the other session the adaptors were the final-/d/ words; order of adaptor (/b/ versus /d/) was counterbalanced across participants. Each session took approximately 20 minutes.

During the initial identification task, listeners heard 18 randomizations of the eight members of the /ba/-/da/ continuum. They identified each syllable by pushing one of the two labeled buttons (“B” on the left key, “D” on the right key). Trials began 500 msec after the previous responses, with a maximum of 3000 msec allowed before moving on to the next trial. During the adaptation task, they made the same judgments, on the same syllables, with the same timing. However, each of the 14 randomizations was preceded by an adaptation phase in which the participants listened to the adapting words without making any responses. During each adaptation phase, the appropriate set of 25 words was randomized twice, yielding 50 adapting tokens before the participants heard a randomization of the /ba/-/da/ syllables and responded by pushing the “B” or “D” buttons. Adaptor words were separated by 300 msec, producing an

adaptation phase of approximately 40 seconds before each randomization of the test syllables.

The data for all experiments can be found on the *Journal's* archive.

Results and Discussion

As in previous studies (e.g., Samuel 1989, 2016), participants who were unwilling or unable to do the task were identified on the basis of their labeling of the /ba/-/da/ syllables. If for either of the adaptation functions the percentage of "D" report for the most /d/-like token was not at least 60% greater than the percentage for the most /b/-like item, the listener was classified as not having done the required task. Two participants in the Original condition and two participants in the No-Release condition were eliminated on this basis, leaving 25 usable participants in the Original condition, and 24 in the No-Release condition.

Following BKA16, the core question is whether listeners identified the test items differently when they were presented after hearing many final-/b/ words than after hearing many final-/d/ words. Figure 2 shows how listeners identified the test syllables in the Original condition as a function of the repeated final consonant in the adaptors. The two notable features in the figure are the clean labeling of the test items (identification as "D" increases smoothly across the continuum), and the small but systematic displacement of the two curves (the red curve is lower than the blue curve in the range where tokens are somewhat ambiguous).

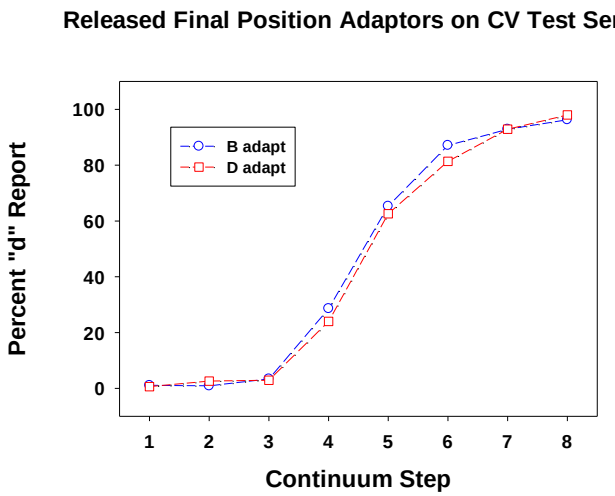

***Figure 2**: Identification of the members of the /ba/-/da/ test series, as a function of the adaptation condition, for adaptors that included released final stops. After adaptation with final-/d/ words (red curve/squares) identification as "D" was reduced compared to adaptation with final-/b/ words (blue curve/circles).*

Recall that the measure of adaptation is based on identification of the test syllables in the ambiguous region of the continuum, defined a priori as items 3-6. For each listener, the average "D" report in this region was computed for the /b/ adaptation case and for the /d/ adaptation case. A simple one-tailed t-test (appropriate given the unambiguous directionality of the test, and the result reported by BKA16) yielded a significant difference, $t(24) = 1.983$, $p < .05$. This result replicates the key finding by BKA16: Stop consonants in word-final position produced a significant shift in identification of initial-position stop consonants.

The results shown in Figure 2 come from the Original condition, using the same adapting words that BKA16 used. The key question in Experiment 1 is whether this result is an artifact of there being released stops in these stimuli. The No-Release condition tests this question by looking at adaptation with identical items, except for the removal of the releases in the final stops. Figure 3 shows the results for this test.

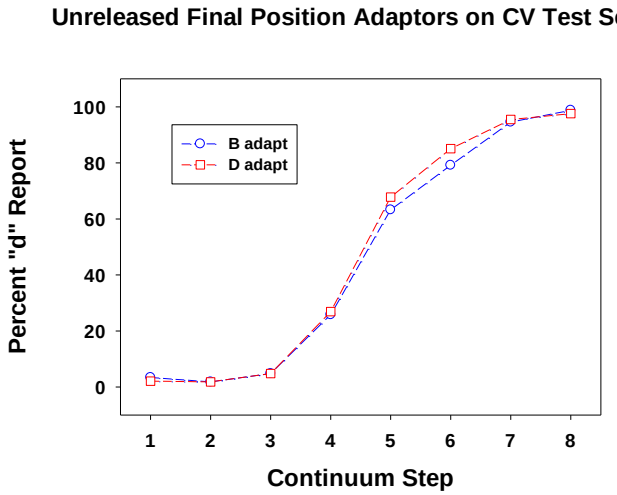

Figure 3: *Identification of the members of the /ba/-/da/ test series, as a function of the adaptation condition, for adaptors that included unreleased final stops. After adaptation with final-/d/ words (red curve/squares) identification as "D" was* ***not*** *reduced compared to adaptation with final-/b/ words (blue curve/circles); there is a small, non-significant reversal.*

As is clear in Figure 3, when the adaptors did not have released final stops, the adaptation effect vanished – the points lie on top of one another for 6 of the 8 steps, and

for the two steps where they separate slightly (steps 5 and 6), the trend is in the wrong direction. This small reversal is clearly not significant, t(23) = -0.869.

The implications of the two conditions are quite clear: When the final-position adaptors did not have extra acoustic information (releases) that matched initial-position stop consonants, no adaptation occurred. When the adaptors from BKA16, including their releases, were used, they did produce a significant shift. These results are completely consistent with the existing adaptation literature (see Table 1). In the absence of matching acoustic cues, adaptation is position-specific; when matching acoustic cues are present (as in Samuel's (1989) experiment with liquids, Wolf's (1978) experiment with stops that included bursts, and in BKA16's study), small but significant cross-position adaptation can be seen. Clearly, this pattern is not supportive of the claim that adaptation is being driven by the kind of abstract position-invariant phonemes developed in linguistic theory.

EXPERIMENT 2

Although the results of Experiment 1 are clear, there are two additional tests that can help put the results in perspective. Both of these situations involve the matched-position cases that have consistently produced adaptation. One situation is a test of initial-position adaptors on the initial-position /ba/-/da/ test series used in Experiment 1. Together with Experiment 1, this test provides a comparison of the results using the more-standard adaptation methods here to the results using BKA16's procedures. Those authors included a condition with initial-position word adaptors on their "bump-

dump” test item, and found that the across-position effect was about one third to one half as large as the matched-position case. The results for Experiment 2’s test using initial-position adaptor words (Initial-Position-Matched) will be used to assess whether the across-position effect shown in Figure 2 is also about one third to one half as large as the matched-position case.

The other test in Experiment 2 is also position-matched, but in this case it is final-position (Final-Position-Matched), rather than initial position. A possible concern about the null effect in Experiment 1 for the No-Release stimuli is that conceivably, in the process of trimming the releases, so much was removed that the remaining word-final information was simply too weak to produce adaptation at all. To assess this, the No-Release adaptors are used on a final-position test series. If the trimming process left sufficient final-position information, then these adaptors should be effective on a final-position test series; if the trimming was excessive, then the adaptors should be ineffective, as they were in Experiment 1.

Method

Participants

A total of 60 participants took part in Experiment 2, 30 in the Initial-Position-Matched situation, and 30 in the Final-Position-Matched case. All were native speakers of American English, with no self-identified hearing problems; none had participated in Experiment 1. They received credit toward a course requirement for their participation.

Stimuli

Test syllables: For the Initial-Matched-Position test, the 8-step /ba/-/da/ test continuum from Experiment 1 was used. For the Final-Matched-Position test, each member of the 8-step continuum was flipped in time, yielding an /ab/-/ad/ test series. This is the same procedure that was used by Samuel (1989) to produce syllable-final stop consonants.

Adaptors: In the Initial-Matched-Position condition, the adaptors were the 25 initial-/b/ and the 25 initial-/d/ words used by BKA16. These had been recorded by the same male speaker of British English who recorded the final-position adaptors (see the original BKA16 paper for a list of the words). In the Final-Matched-Position case, the No-Release adaptors (i.e., the trimmed versions) from Experiment 1 were the adaptors.

Apparatus and Procedure

Participants were tested in the same sound-shielded chambers, with the same apparatus and procedures as in Experiment 1.

Results and Discussion

The same criteria were used to identify participants who did not do the task as instructed. Five participants in the Initial-Matched-Position condition and seven participants in the Final-Matched-Position condition were eliminated on this basis, leaving 25 usable participants in the first case, and 23 in the other.

Figure 4 shows the adaptation results for the Initial-Matched-Position test. Consistent with the prior adaptation literature (see Table 1), adaptors that match test syllables in position produce reliable shifts, $t(24) = 2.415$, $p < .05$. Comparing Figure 4

(matched-position) to Figure 2 (across-position), it is clear that the effect was larger for the matched-position case. More specifically, using the average identification of the middle four items of the test series, there was a shift of 8.3% in Experiment 2, versus a shift of 3.4% in Experiment 1. Thus, the size of the across-position effect was 41% of the size of the matched-position effect. Using these same adaptors, with their "bump-dump" test item, BKA16 reported an across-position effect that was 34% of the shift for their matched-position case. When they only considered the subset of listeners without floor or ceiling effects, this value was 48%. These values exactly bracket the value in the current study.

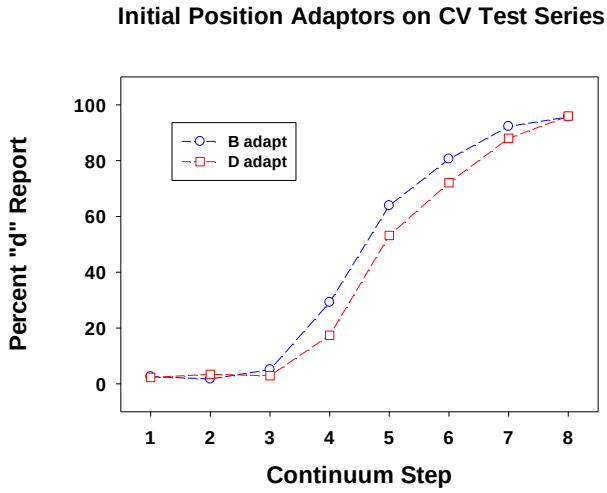

Figure 4: *Identification of the members of the /ba/-/da/ test series, as a function of the adaptation condition, for adaptors that included initial stops. After adaptation with initial-/d/ words (red curve/squares) identification as "D" was significantly reduced compared to adaptation with initial-/b/ words (blue curve/circles).*

Figure 5 shows the results for the Final-Position-Matched adaptation test. Recall that in this case the adaptors were the trimmed No-Release final-position words, and the test items were vowel-consonant (VC) syllables that were mirror images (in the time domain) of the consonant-vowel test items used in all of the other cases. The purpose of this test was to determine whether the process of trimming the releases from BKA16's adaptors had removed too much of the final consonantal information. If that were the case, the failure of the trimmed adaptors in Experiment 1 would not be informative. As is clear in Figure 5, the trimmed adaptors retained the critical final consonantal information: These adaptors produced a very robust shift in the identification of the final position stops of the /ab/-/ad/ test items, $t(22) = 7.150$, $p < .05$. Thus, the failure of these same adaptors to produce an effect on the initial-position test items in Experiment 1 is not an artifact of the trimming process. Rather, it is another example, like those in Table 1, of the ineffectiveness of across-position adaptors.

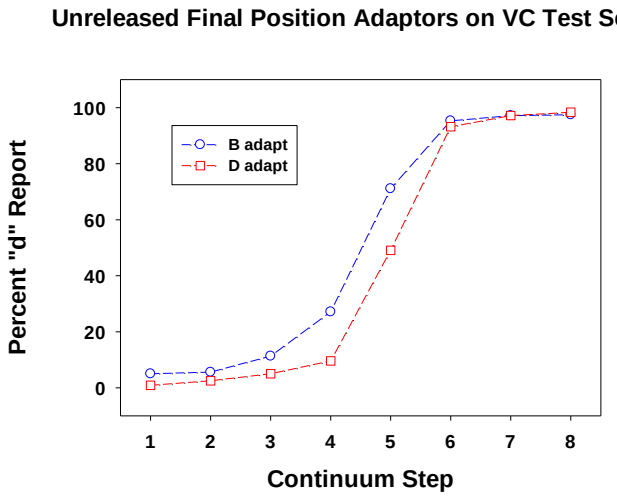

***Figure 5**: Identification of the members of the /ab/-/ad/ test series, as a function of the adaptation condition, for adaptors that included unreleased final stops. After adaptation with final-/d/ words (red curve/squares) identification as "D" was reduced compared to adaptation with final-/b/ words (blue curve/circles).*

GENERAL DISCUSSION

Recall that the current study is intended to address three issues: (1) *Empirically*, do the data reported by BKA16 support their advocacy of position-invariant phonemes as perceptual units? (2) *Methodologically*, is the resurgence of the selective adaptation paradigm well-informed? and (3) *Theoretically*, should psycholinguistic investigations of perception rely on units derived through linguistic analysis? I will consider each of these in turn.

The results of the two experiments here provide a clear "No" to the empirical question: The apparent position-invariant adaptation effect reported by BKA16 is an artifact of the release bursts in their final-position adapting words. Instead of providing evidence at odds with the prior literature, their experiments (and the experiments here) add to the already substantial literature summarized in Table 1 – there are now additional positive results for position-matched adaptation, negative results for position-mismatched adaptation, and additional examples of successful adaptation driven by acoustic matching.

The empirical aspect of the current study was entirely focused on the BKA16 paper. For the methodological and theoretical issues, it is important to consider the

MRM18 paper as well, beginning with a brief summary of its goals, methods, and conclusions. With minor exceptions, the methods were quite similar to those of BKA16, though they did avoid the use of a single ambiguous token (like the “bump-dump” token) by testing a few tokens near the middle of a continuum. This brought their methods more in line with the existing adaptation literature.

Their goals and conclusions were stated clearly at the beginning and end of their paper, respectively. Their goal was to answer the “units” question: “In spoken-word recognition, the question is which code we use to map the highly variable speech signal onto knowledge stored in the mental lexicon... What, in short, are the pre-lexical units of speech perception?” (p. 77). Throughout the paper, the answer they gave to this question is the allophone. For example, they said “At a theoretical level, it is important to emphasize that, irrespective of whether the adaptation process itself concerns changes in the representations of allophones, changes of the auditory patterns that define those allophones, or changes in the mapping between these two types of representations, the present data indicate that it is knowledge about allophones, not about phonemes, that is involved.” (p. 89) Their concluding paragraph left no ambiguity about the theoretical claim: “We have provided evidence that context-insensitive phonemes are not a part of spoken-word recognition. … The present findings, along with other recent data using the perceptual-learning paradigm, suggest instead that pre-lexical processing is based on allophones. This proposal has clear implications for models of speech recognition. As summarised in the Introduction, most models assume the pre-lexical units of speech perception are not allophones; in this regard, these models may all be incorrect.” (pp. 90-91). It is important to see that with respect to the units question, MRM18 had two

core goals/claims: First, to show that abstract phonemes (as advocated by BKA16) are not perceptual units, and second, to show that allophones are.

Note that MRM18 assert that the field has not embraced allophones ("most models assume the pre-lexical units of speech perception are not allophones"). This is notable because essentially the same complaint is made about the phoneme by Bowers and his colleagues: "[T]here is a long history of challenging the phoneme hypothesis, with some theorists arguing for differently sized phonological units (e.g. features or syllables) and others rejecting abstract codes in favour of representations that encode detailed acoustic properties of the stimulus. The phoneme hypothesis is the minority view today." (Kazanina, Bowers, & Idsardi, 2018, p. 560). Given the claim that phonemes are "the minority view", Bowers and his colleagues set out to demonstrate that there is *some* perceptual role for phonemes. In the absence of any thorough review of the literature, neither claim of disrespect is based on any quantitative analysis of how popular either unit actually is. My own impression is that the phoneme is widely assumed by most people in the field, but importantly, this assumption is not of a formal phonemic unit. Rather, when people invoke phonemes, they are simply referring to the idea of vowels and consonants at a surface level. For example, in my demonstrations that lexical activation can support perception of phonemes (e.g., Samuel 1997, 2001), my point was that if a listener hears "aboli?", with the "?" representing an ambiguous mixture of "s" and "sh", the listener will perceive the consonant "sh" (whereas the percept will be of "s" if the same "?" mixture occurs in "malpracti?"). There was no intent to be precise about whether that perceived consonant should be thought of as a phoneme or allophone (or some other linguistic unit) – it was simply a consonant (which

psycholinguists typically consider to be a phoneme, but not in a linguistic, formal, sense).

In the current context, we need to be more precise. Recall that in linguistic theory, both allophones and phonemes are abstractions, with phonemes being larger sets that can include multiple allophones (e.g., the phoneme /p/ includes both the aspirated allophone [p^{h}] that occurs in initial position, and the unaspirated allophone [p] that occurs in non-initial position). Thus, although the MRM18 paper was designed as a direct challenge to BKA16, both sets of authors embraced the assumption that the unit being sought was to be found among those that linguists postulated; their disagreement is about *which* linguistic unit is key.

In order to disprove phonemes as "the pre-lexical units of speech perception" (p. 91), MRM18 chose two test domains that involved cases in which linguistic theory produces particularly abstract relationships between phonemes and the actual sounds that appear in the speech stream. Their goal was to show that there is no adaptation between adaptors and test items of this sort. The argument is that if phonemes are truly the pre-lexical units, then adaptation should occur despite the big acoustic differences. For example, in Dutch (one testing domain), the phonemes /r/ and /l/ have different surface forms in initial versus final position. The "light" form of /l/ appears in initial position, and the "dark" form in final position; there are even larger differences in the phonetic realization of /r/ across positions. If both forms get mapped onto the underlying phoneme, and phonemes are perceptual units, then adaptation should occur despite the quite different acoustic patterns in a "light" adaptor and a "dark" test item (or vice versa). The second test domain was even more challenging for an abstract phoneme model, as

the surface forms for certain German fricatives are very different acoustically. In both cases, there was little or no adaptation found when adaptors and test items were acoustically very different, despite sharing a common phoneme in linguistic theory. Thus, MRM18 rejected the phoneme as the pre-lexical perceptual unit. This rejection of abstract (linguistic) phonemes is well supported by their results. Critically, however, MRM18 went a step further, and argued that their findings instead supported allophones (a different abstract, linguistic unit).

MRM18 and BKA16 thus engaged in exactly the type of endeavor that Goldinger and Azuma (2003) warned against. Before considering this further, it is worth noting that both sets of authors conducted their tests using the adaptation technique, but that neither drew on the existing adaptation literature to inform their undertaking. As Table 1 summarized, the positional issue that BKA16 examined was previously tested in five prior studies, comprising 18 within-position tests and 18 across-position tests. Similarly, MRM18's test of liquids with varying allophones had also already been reported in the adaptation literature. Like Dutch, English has both "light" and "dark" versions of /l/ (Sproat & Fujimura, 1993), and different phonetic realizations of /r/ across position. In the same paper that was a basis for the BKA16 study, Samuel (1989) tested whether adaptation occurs across the phonetic variants of /r/ and /l/. The framing of the experiment was more in terms of position, but it involved the same test of light versus dark liquids, and aside from the effects based on the steady-state acoustic overlap (see Table 1), the experiment produced the same null effect that MRM18 found 30 years later.

At a theoretical level, it is worth noting that the units issue has been a focus of study in the adaptation literature, and that there are several findings that seem problematic for a theory that favors allophones, phonemes, or any particular linguistic units. For example, Ganong (1978) tested whether adaptors with burst-cued stop consonants could produce adaptation on a stop-consonant place of articulation test series produced with only formant transition cues; they did. In fact, burst-cued stops could cause shifts on a nasal (/m/-/n/) place of articulation test series, even though the nasals are never cued by bursts. Samuel and Newport (1979) reported a similar effect of adaptation being driven by one cue in an adaptor (the gradual onset of /ʃ/ ("sh") versus the abrupt onset of /tʃ/ ("ch")) on a test series that differed along a different acoustic dimension (longer duration of the frication for /ʃ/ than for /tʃ/). The results of these studies are difficult to reconcile with an allophone-based model. Similarly, any theory that relies on linguistic units (whether allophones or phonemes) will have difficulty accounting for Diehl's (1976) finding that a nonspeech musical sound could shift the boundary on a /b/-/w/ test series (cf. Kat & Samuel, 1984, and Samuel & Newport, 1979, for replications and extensions of these findings).

There is another subset of the adaptation literature that is germane to the units issue, and that again is problematic for models based on allophones or phonemes. The studies in this subset employed an approach called "contingent" adaptation. In one such paper, Sawusch and Pisoni (1978) tested adaptation by pairs of alternating adaptors (e.g., /ba/ - /di/) on test series with different vowels (e.g., a /ba/-/da/ test series, and a /bi/-di/ test series). Adaptation was "contingent" – the /ba/-/di/ adaptors shifted /ba/-/da/ as it would be shifted by /ba/, but the /ba/-/di/ adaptors shifted /bi/-/di/ as it

would be shifted by /di/. Such contingent adaptation seems to support consonant-vowel units (additional support for such units, using non-adaptation techniques, comes from Sumner & Samuel, 2007), rather than either phonemes or allophones (note that consonant-vowel units are not typical units in linguistic theory, as phonemes or allophones are).

The papers mentioned here are just a few examples of results in the adaptation literature that are germane to the perceptual units issue. As Figure 1 illustrates, there is a large body of research that employed the adaptation paradigm. Many of these early studies examined issues (such as the units issue) that continue to be of interest to researchers. Adaptation can be an effective way to investigate a number of important current theoretical issues, as the recent increase in its use indicates. Researchers should be informed by what has already been established in the extensive prior literature, but many recent adaptation studies fail to draw upon these findings. Studies of how listeners recalibrate their phoneme boundaries when exposed to ambiguous phonetic input often involve stimulus presentation conditions that are formally similar to those in an adaptation test, but these studies very rarely even acknowledge this, let alone draw on the adaptation literature.

Poor awareness of older research is of course not a phenomenon that only is found in current adaptation and recalibration research. For example, the concept of "prediction" and the brain's use of "prediction error" are enjoying a boom in research, in part driven by imaging techniques that allow researchers to see brain activation patterns that precede observable behavior. Many of the issues that are being examined in this domain are ones that were studied previously under the rubric of top-down processing –

situations in which a person uses partial information to construct the likely remainder of an input. Yet, much of the prediction literature fails to engage with the earlier work (perhaps because of the current focus on imaging evidence, which was not available in the earlier literature). To a certain extent, the situation for adaptation and recalibration is exacerbated by the long period during which very little adaptation research was published (see Figure 1). Thus, even though the problem is quite general, it may be particularly acute in these areas of research.

Clarifying an erroneous finding in the literature (the across-position effect that was due to a stimulus artifact) is important to prevent other researches from expending effort unproductively; this is uncontroversial. Similarly, researchers would all presumably agree that their work should be informed by relevant prior research. At this point, I will turn to a potentially more controversial position: Psycholinguists have expended too much effort testing the psychological reality of units that have been created by linguists. To be clear, linguistic theory has played an extremely important role in cognitive science, and it continues to do so. Moreover, as I will discuss below, linguistic units can serve important functions. Thus, my argument here should in no way be seen as a criticism of linguistics. My criticism is instead of psycholinguists: Things can go awry when psycholinguists treat the *descriptive* structures in linguistic theory as processes in decoding speech input.

Perhaps the most influential descriptive structures in linguistic theory were the transformations that Chomsky (1957, 1965) posited in his profoundly important syntactic theories. Chomsky described transformations that could apply to a “deep structure” in order to produce the “surface structure” – the actual sequence of words in a sentence.

Psycholinguists latched onto the idea of transformations and launched a very substantial research effort to prove the “Derivational Theory of Complexity”: The processing difficulty for a listener should be a function of the number of transformations that would be needed to develop the surface structure from the deep structure. In a brilliant book, Fodor, Bever, and Garrett (1974) described the many elegant experiments that were inspired by the Derivational Theory of Complexity. Ultimately, after years of these experiments, it became clear that even if transformations were valuable linguistic constructs, they did not actually explain language *processing*. We might dismiss this as an error made when the field was young and inexperienced, but that would neglect the current boom in studies looking to relate syntactic structures to fMRI activation patterns (e.g., Brennan, Stabler, van Wagenen, Luh, & Hale, 2016).

In the domain of speech perception, the adoption of linguistic units as processing units is similarly common. BKA16 framed their study as a response to what they perceived to be a widespread rejection of the phoneme as a perceptual unit: “[W]e review the current empirical evidence regarding phonemes in the domains of speech production and perception, and then describe two experiments that provide strong evidence that phonemes do indeed play a role in word perception.” (p. 72). In over a dozen places in the paper they advocate for the phoneme, with sentences like “We take these findings to support the claim that position independent phoneme representations play a role in speech perception” (p. 79), and “Our findings are also consistent with a number of number of speech perception studies that have provided evidence for phonemes” (p. 80). Occasionally, their advocacy for the phoneme is presented as a contrast to the allophone: “Toscano, Anderson, and McMurray (2013) provided

evidence that phonemes are coded independently of position... Note that the effect cannot be explained at the allophone level" (p. 74).

In a follow-up paper (Kazanina, Bowers, & Idsardi, 2018), the primacy of the phoneme is pushed a bit harder, with the view that phonemes are the units in the mental lexicon and thus the necessary entry codes for lexical access: "[P]honemes are access codes to the lexicon (i.e., the sublexical representations retrievable from the acoustic signal that directly interface with phonological forms of words)." (p. 562). In both papers, the authors are unambiguous in drawing upon linguistic analysis for their units. In fact, the linguistic perspective is dominant, with multiple statements of the following sort: "It is the linguistic arguments that provide the strongest evidence for the psychological reality of phonemes as access units in speech perception that can support further language comprehension." (p. 561).

Across the two papers, Bowers and his colleagues thus make the relatively mild assertion that phonemes "play a role" in speech perception, but also make it clear that phonemes have a uniquely important role because they are both the internal representations in the lexicon and the access codes to the lexicon during speech perception. For the moment, I will focus on the milder claim that phonemes are just one of several perceptual units. In assessing this claim, a central question is whether the term "phoneme" is specifically intended to be the abstract unit that linguists have defined. Both papers seem to be making this specific claim. If so, there is a fundamental problem: The empirical results demonstrate that the position is incorrect. The data in the current study show that when the test chosen by BKA16 is run with proper stimuli, no evidence is found for abstract phonemes. Moreover, when MRM18 designed two

additional tests that follow from the view that abstract phonemes play a perceptual role, the results were again counter to the theory.

A perceptual role for phonemes could be salvaged by stepping back from the abstract linguistic framing of the unit. If “phoneme” were instead to be viewed as simply a vowel or consonant of the language, as the term is typically used in psycholinguistic research, then it would be easy to identify many uses in the literature. I have already mentioned one such use – the idea that lexical context can generate a “phoneme” from noise (Samuel, 1997) or from an ambiguous segment (Samuel, 2001). The most commonly invoked models of speech perception and word recognition (e.g., Elman & McClelland’s (1986) TRACE model, and various models by Norris and his colleagues, e.g., Norris, McQueen, & Cutler, 2000) have phoneme-like units. There are literatures that assume that listeners have access to phonemes, including work on phonemic restoration (e.g., Warren, 1970). One of the most popular paradigms in speech research for many years was the phoneme monitoring task, in which listeners are given a target (e.g., the sound “b”) and told to push a button when they detect an occurrence of that phoneme. Critically, none of these uses in the literature are grounded in the linguistic/abstract interpretation of a phoneme. If BKA16 were not talking about this linguistic idea of a phoneme, there would be ample evidence for the unit playing a role in perception, but their presentation makes it clear that they do mean the linguistic/abstract version.

MRM18 are somewhat less explicit about the linguistic basis for their position, but they leave no doubt that their goal is to advocate for the allophone as the key perceptual unit. There are about two dozen statements in the paper that promote the

allophone, downplay the phoneme, or in most cases, both. For example, “If listeners have allophonic units, they could optimize the mapping of the input onto the lexicon for each allophone separately. This would be harder to achieve with phonemic units” (p. 78). Or, “[P]honemic identity has little role to play in functional adaptations in speech perception” (p. 79). Or, “[T]he present data indicate that it is knowledge about allophones, not about phonemes, that is involved.” (p. 89). Thus, the purpose of BKA16’s paper was to provide evidence for the psychological reality of one linguistic unit (the phoneme), while the goal of MRM18 was to instead show that a different linguistic unit (the allophone) is psychologically real.

Attempts to provide psycholinguistic support for linguistic units have been widespread in our field. For example, Kraljic and Samuel (2006) investigated generalization of lexically-driven perceptual recalibration, and argued that such generalization was well-explained if the recalibration was grounded in learning at the level of the phonetic feature (a linguistic unit). The evidence for this claim was that listeners who had undergone recalibration on a /d/-/t/ (voicing) distinction showed just as much recalibration when they were tested on their perception of a /b/-/p/ (voicing) distinction. Thus, the pattern was neatly accounted for if listeners had made adjustments to how the voiced-voiceless feature was specified.

Although an account relying on phonetic features was elegant, it was also wrong. Schuhmann (2014) used a similar recalibration design to look for generalization across phonetic features, and showed that the pattern is not as simple as Kraljic and Samuel (2006) had suggested. Her listeners underwent recalibration on the distinction between /f/ and /s/, and showed generalization to a /v/ - /z/ contrast, but not to a /p/ - /t/ contrast.

All three of these contrasts involve the same labial (or labial-dental) versus alveolar place distinction, and there was generalization across voicing, but not across manner (fricative versus stop). In linguistic theory, there is no obvious reason why one feature (voicing) would allow generalization, while another (manner) would not. These results suggest that the account that Kraljic and Samuel offered, in terms of a phonetic feature, was misguided.

This example is instructive because it is typical of efforts to invoke linguistic units in psycholinguistic investigations: Initially the explanation works well, but as more information accumulates, the account breaks down. At the beginning of their paper, MRM18 make this point nicely. After noting that their interest is in determining the pre-lexical units of speech perception, they provide a list of some of the units that have been suggested – abstract phonological features, context-dependent allophones, context-independent phonemes, and syllables. They then accurately describe the problem: “One recurring issue in this long-running debate has been that evidence in favour of one or the other type of unit often turned out to be paradigm-specific. Evidence for many different units can therefore be found (for a review, see Goldinger & Azuma, 2003).” (p. 77).

However, rather than accept their own warning, they suggest that this time, things will work out: “Recent evidence from learning and adaptation paradigms has breathed new life into this debate. This is because such paradigms offer the possibility of establishing which units play a role in speech perception by asking which units are learned about, and thus offer a more direct measure than the classic paradigms“ (p. 78). Unfortunately, in my opinion, this optimism is groundless – there is no reason to believe

that learning (recalibration) and adaptation are any more likely to reveal true perceptual units than other paradigms, regardless of the many other useful things that these paradigms can tell us. MRM18 are particularly hopeful about the recalibration paradigm's potential because "it is based on processes which are involved in solving the critical problem in spoken-word recognition, the invariance problem. The paradigm thus reveals units that are functional in speech perception. That is, it reveals units that are involved in active adaptation to variance in the input and that hence help the listener decode the highly variable speech signal." (p 89). To the best of my knowledge, there is currently no clear evidence that recalibration solves the invariance problem. Even if it did, that would not provide any assurance that the task offers any particularly direct window into perceptual units.

In fact, the existing literature already provides ample evidence against recalibration operating at any particular linguistic level. Despite the initial pattern of Kraljic and Samuel (2006) results, Schuhmann's (2014) findings demonstrate that the effects do not support the established linguistic unit of a phonetic feature. Mitterer, Reinisch, and their colleagues have reported recalibration results that are similarly problematic for linguistically-defined phonemes: Mitterer and Reinisch (2013) conducted a recalibration study with the same liquids used in MRM18's adaptation test, and found a comparable lack of transfer across allophonic variation. Similarly, Mitterer and Reinisch (2017) found no transfer from devoiced to voiced stops, even though in linguistic theory these are phonological variants of each other. And, Reinisch, Wozny, Mitterer, and Holt (2014) reported that recalibration can be vowel-specific, with learning

about [aba] vs [ada] failing to generalize to [ibi] vs [idi]. This result is reminiscent of the contingent adaptation effects described above (e.g., Sawusch & Pisoni, 1978).

Despite their goal of advancing the allophonic position, MRM18 state that “what matters in perceptual learning is auditory overlap rather than abstract featural overlap” (p. 90). Note that auditory overlap is not the same as allophonic overlap – by definition, “auditory” is not linguistic, and the allophone is a linguistic construct. Kraljic and Samuel (2007) found that recalibration for fricatives was speaker-specific, but recalibration for stops was speaker-general. Again, there are no linguistic units that can accommodate this pattern. Other findings in the recalibration literature are similarly not well-matched to any particular linguistic units. It seems likely that as the recalibration literature grows to the size of the adaptation literature, there will be evidence for effects at multiple levels, just as has been found in adaptation (e.g., Samuel & Kat, 1996; Sawusch, 1977a).

Despite their fundamental goal of advocating for one linguistic unit or another, BKA16 and MRM18 seem to have been aware of these kinds of complications. Thus, each paper includes a statement that contrasts with the dozens of statements that are made to support the preferred unit. BKA16 say that “[N]o one claims that phonemes are the sub-lexical unit of perception (that is, the only sub-lexical unit). Rather, the claim is that phonemes are a sub-lexical unit of perception (that is, one of perhaps several sub-lexical units).” (p. 74). MRM18 say “[A]llophones are not the only type of abstraction that supports generalization of learning. … Some structures may be smaller than segments, such as aspiration or the release bursts as parts of voiceless stops. ... Other structures may be larger than a segment. … What may matter for perceptual learning is

not the grainsize of the structure per se, but rather whether the structure is a consistent production pattern in the interlocutor's speech." (p. 90). It is difficult to reconcile these statements with the core message of each paper, as the core message of BKA16 is that the phoneme is a key perceptual unit, whereas the core message of MRM18 is that the allophone is the key perceptual unit. The similarity of the positions expressed in the two quotations here is notable, given that the repeated claims in the two papers are diametrically opposed.

It is instructive to compare MRM18's statement to the perspective on units that Grossberg, Boardman, and Cohen (1997) offered in the context of the Adaptive Resonance Theory (ART): "The language units that are familiar to us from daily experience, such as phonemes, letters, and words, do not form appropriate levels in a language processing hierarchy… Rather, processing levels that compute more abstract properties of auditory processing are needed; in particular, a working memory… is posited herein that represents sequences of 'items' that have been unitized through prior learning experiences. Such items are familiar feature clusters that are presented within a brief time period… These postulates lead to working memories that can store sequences of events in a way that enables them to be grouped, or unitized, into categories, or 'list chunks'…These list chunks may represent the items themselves or larger groupings of items, such as phonemes, letters, syllables, or words." (p. 482). MRM18 suggest that a unit is a "structure [that] is a consistent production pattern in the interlocutor's speech", which seems virtually identical to ART's notion of items "that have been unitized through prior learning experiences". Similarly, MRM18's statement "Some structures may be smaller than segments, such as aspiration or the release

bursts as parts of voiceless stops. ... Other structures may be larger than a segment" appears to be entirely consistent with the ART position that items can be unitized into "list chunks" that "may represent the items themselves or larger groupings of items, such as phonemes, letters, syllables, or words". To the extent that BKA16 are only saying that the phoneme is one of a number of perceptual units, their position also aligns with ART. However, critically, their advocacy is for an abstract phoneme, and the units in ART are based on patterns that are repeatedly encountered in the signal.

The futility of trying to reify linguistic units is illustrated by the convergence of the two positions, and by their close mapping onto a theory that explicitly rejects the attempt to build a psycholinguistic model based on linguistic units. In fact, in ART, it is just the reverse: The items and list chunks in perception mimic linguistic units to the extent that such units are generally reflected in the probabilistic exposure pattern for a listener; when high frequency patterns occur that are not standard linguistic units (such as the CV patterns implicated in the contingent adaptation literature), ART does not distinguish between these non-linguistic units and ones that coincide with the units linguists have developed to describe language. In fact, after noting the failure of the field to converge on any particular perceptual unit in speech perception, Goldinger and Azuma (2003) recommended ART's approach. As they put it, "self-organization through adaptive resonance … simply nullifies the ‘‘units’’ question." (p. 307). It is worth noting that ART would also predict that linguistic units that do not correspond to frequent input patterns will not be viable processing units, consistent with MRM18's null adaptation findings for adaptation based on abstract phoneme units that have little acoustic overlap.

I believe that the evidence is very clear that linguistic units, such as phonemes or allophones, do not have any privileged status in the process of spoken word recognition. As Grossberg et al. (1997) suggest, the perceptual system is omnivorous – if the input consistently includes a particular pattern, that pattern can be learned as a "chunk", and such chunks will be used to recognize speech. This will be true whether the common pattern corresponds to a linguistically-defined unit, or to some configuration that does not play a role in linguistic theory. That is why, if one looks broadly at the adaptation literature, or at the recalibration literature, there are results that align with use of both linguistically-defined and non-linguistically-defined units.

The perceptual system's agnostic treatment of units does not undercut the potential importance of linguistic units in other ways. Together with BKA16, Kazanina, Bowers, and Idsardi (2018) review a large number of linguistic and psycholinguistic phenomena that are best understood in terms of linguistic units (especially phonemes). Along similar lines, many aspects of speech production are most coherently described by assuming underlying (abstract, linguistic) syllables, even though the evidence for a perceptual role of the syllable in English is extremely weak (there is better evidence in French, consistent with syllables being better defined in the acoustic signal in French). Thus, the argument against phonemes and allophones as privileged perceptual units should not be taken as an argument against phonemes and allophones across the board. In fact, even within perception, there will be situations in which something approximating a phoneme or an allophone will be a perceptual unit, simply because the evidence in the input can promote the development of a chunk of that sort. The

argument made here is that in the domain of speech perception, assuming a privileged status for linguistically-defined units is unjustified, both theoretically and empirically.

In the Introduction, I noted that Goldinger and Azuma (2003) had already highlighted the futility of looking for a particular linguistic unit in perception, based on what at that point was 30 years of unsuccessful attempts. If 30 years of failure were not enough to discourage us from pursuing a fruitless agenda, my argument is that 50 years should be. Almost a half-century ago, McNeill and Lindig (1973) presciently said "What is "perceptually real" is what one pays attention to. In normal language use the focus of attention…is the meaning of the utterance. Subordinate levels become the focus of attention only under special circumstances. The normal perceptual object in speech, the focus of attention, therefore, is *none* of the linguistic levels that have been studied in monitoring experiments…[T]here is no clear sense in which one can ask what the "unit" of speech perception is." (p. 430, emphasis in the original).

It is impressive that within less than a decade of relevant research, McNeill and Lindig (1973) were able to identify the futility of testing the perceptual reality of "linguistic levels". Unfortunately, the field did not listen to their warning, leading to a substantial waste of valuable research effort on misguided undertakings such as trying to prove the Derivational Theory of Complexity in the 1970's. As Goldinger and Azuma (2003) noted, these efforts continued into the 1980's and 1990's, leading to their warning about the "quixotic quest" for linguistic units in speech perception. In the two decades following their paper, this quest has continued, including the two recent papers that have been the focus here.

If researchers accept the argument that these attempts are misguided, does that mean that we should stop trying to understand how speech is encoded? Of course not. There are many fundamental questions about encoding that are, and should be, a focus of research. For example, the question of position-specific versus position-general sensitivity that BKA16 tested strikes me as a very appropriate issue in understanding how the speech system sorts the input. It is the framing/motivation of the research in terms of linguistic units that is problematic. Given that the system will latch onto virtually any systematic pattern in the input, there is no news in finding that a particular pattern is used. The news is that the system does treat position as a relevant dimension. Similarly, finding that the speech system sorts the input in terms of the periodicity of the sounds, and their onset properties (Kat & Samuel, 1984; Samuel & Newport, 1979) helps to understand the encoding process, without preemptively excluding properties that are non-linguistic. Studying how the encoding changes as a function of phonetic ambiguity, with a teaching signal provided by visual speech (Bertelson, Vroomen, & van Gelder, 2003) or by lexical constraints (Norris, McQueen, & Cutler, 2003) has provided a wealth of information about speech processing, without any need to tie to research to units that linguists have developed to describe speech.

The list of potentially productive research programs could go on and on – it is virtually infinite. The argument that I have made does not impose daunting constraints on possible research programs in the domain of speech perception -- there is absolutely no problem finding interesting and productive research questions to pursue. The constraint that I have suggested is simple, even though it goes against a tempting impulse: Because the system uses whatever patterns the input provides, whether these

correspond to linguistic units or not, researchers should not undertake an effort to simply show that some linguistic unit plays a role in speech perception. Linguistic analyses can provide very useful ways for researchers to think about speech, but psycholinguists should not assume that linguistic units have a privileged status in perceptual processing.

Acknowledgments

Support provided by Economic and Social Research Council (UK) Grant #ES/R006288/1, Ministerio de Ciencia E Innovacion (Spain) Grant # PSI2017-82563-P and by Ayuda Centro de Excelencia Severo Ochoa (Spain) SEV-2015-0490.

The data for all experiments can be found on the *Journal's* archive.

Jeff Bowers provided me with the stimuli used in Bowers, Kazanina, and Andermane (2016). I sincerely appreciate his providing the stimuli used in this project. Similarly, I thank Holger Mitterer for providing me with constructive feedback on an earlier version of this paper. Finally, I wish to thank Steve Goldinger for his insightful, thoughtful, and constructive review of a previous version of this paper.

REFERENCES

Ades A. (1974). How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, *16*, 61-66.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk after effect. *Psychological Science*, *14*, 592-597.

Bowers, J.S., Kazanina, N., & Andermane, N. (2016). Spoken word identification involves accessing position invariant phoneme representations. *Journal of Memory and Language*, *87*, 71-83.

Brennan, J.R., Stabler, E.P., van Wagenen, S.E., Luh, W-M., & Hale, J.T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, *157-158*, 81-94.

Chomsky, N. (1957). *Syntactic Structures*. The Hague, Moulton.

Chomsky, N. (1965). *Aspects of a Theory of Syntax*. Cambridge, MA: MIT Press.

Diehl, R. (1976). Feature analyzers for the phonetic dimension stop vs. continuant. *Perception & Psychophysics*,*19*, 267-272.

Diehl, R. (1981). Feature detectors for speech: A critical reappraisal. *Psychological Bulletin*, *89*, 1-18.

Diehl R., Elman J., McCusker S. (1978). Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance*, *4(4)*, 599-609.

Eimas, P.D. & Corbit, J.D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*, 99-109.

Elman, J.L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33(4)*, 547-582.

Fodor, J.A., Bever, T.G., & Garrett, M.F. (1974). *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw-Hill.

Foss, D.J. & Swinney, D.A. (1973). On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior*, *12*, 246-257.

Gangong W. F. (1978). The selective adaptation effects of burst-cued stops. *Perception, & Psychophysics*, *24*, 71-83.

Goldinger, S.D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, *31*, 305-320.

Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 483–503.

Kat, D., and Samuel, A. G. (1984). More adaptation of speech by nonspeech. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 512-525.

Kazanina, N., Bowers, J.S., & Idsardi, W. (2018). Phonemes: Lexical access and beyond. *Psychonomic Bulletin and Review*, *25*, 560-585.

Kleinschmidt, D. & Jaeger, T. F. (2015). Robust speech perception: Recognizing the familiar, generalizing to the similar, and adapting to the novel. *Psychological Review*, *122*, 148-203.

Kraljic, T., & Samuel, A.G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, *13*, 262-268.

Kraljic, T., & Samuel, A.G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*, 1-15.

McNeill, D., & Lindig, D. (1973). The perceptual reality of phonemes, syllables, words, and sequences. *Journal of Verbal Learning and Verbal Behavior*, *12*, 419-430.

Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, *69*, 527–545.

Mitterer, H., & Reinisch, E. (2017). Surface forms trump underlying representations in functional generalisations in speech perception: The case of German devoiced stops. *Language, Cognition and Neuroscience*, *32*, 1133–1147.

Mitterer, H., Reinisch, E., & McQueen, J.M. (2018). Allophones, not phonemes, in spoken-word recognition. *Journal of Memory and Language*, *98*, 77-92.

Norris, D., McQueen, J.M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299-370.

Norris, D., McQueen, J.M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204-238.

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category

recalibration: What are the categories? *Journal of Phonetics*, *45*, 91–105

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology*, *18*, 452-499.

Samuel, A. G. (1989). Insights from a failure of selective adaptation: Syllable-initial and syllable-final consonants are different. *Perception & Psychophysics*, *45*, 485-493.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, *32*, 97-127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, *12*, 348-351.

Samuel, A.G. (2016). Lexical representations are malleable for about one second: Evidence for the non-automaticity of perceptual recalibration. *Cognitive Psychology*, *88*, 88-114.

Samuel, A. G., and Kat, D. (1996). Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 676-694.

Samuel, A. G., Kat, D., and Tartter, V. C. (1984). Which syllable does an intervocalic stop belong to? A selective adaptation study. *Journal of the Acoustical Society of America*, *76*, 1652-1663.

Samuel, A. G. and Newport, E. L. (1979). Adaptation of speech by non-speech: Evidence for complex acoustic cue detectors. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 563-578.

Savin, H.B., & Bever, T.G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 9, 295-302.

Sawusch J. (1977a). Peripheral and central processes in selective adaptation of place of articulation in stop consonants *Journal of the Acoustical Society of America*, *62*, 738-750.

Sawusch, J. (1977b). Processing of place information in stop consonants. *Perception & Psychophysics*, *22*, 417-426.

Sawusch, J., & Pisoni, D. (1978). Simple and contingent adaptation effects for place of articulation in stop consonants. *Perception & Psychophysics*, *23*, 125-131.

Schuhmann, K. S. (2014). *Perceptual learning in second language learners*. Stony Brook, NY: State University of New York dissertation.

Sproat, R., & Fujumura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*, 291-311.

Sumner, M., & Samuel, A.G. (2007). Lexical inhibition and sublexical facilitation are surprisingly long lasting. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*, 769-790.

Toscano, J. C., Anderson, N. D., & McMurray, B. (2013). Reconsidering the role of temporal order in spoken word recognition. *Psychonomic Bulletin & Review*, *20(5)*, 981–987.

Vroomen J, van Linden S., de Gelder B, Bertelson P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia*., *45*, 572–577.

Vroomen J, van Linden S., Keetels M, de Gelder B, Bertelson P. (2004). Selective adaptation and recalibration of auditory speech by lipread information: Dissipation. *Speech Communication*, *44*, 55–61.

Warren. R. (1970). Perceptual restoration of missing speech sounds. *Science, 167***,** 392–393.

Wolf C. (1978). Perceptual invariance for stop consonants in different positions. *Perception & Psychophysics*, *24*, 315-326.