



# Effects of lexical ambiguity, frequency, and acoustic details in auditory perception

Chelsea Sanker<sup>1</sup>

Published online: 15 October 2018  
© The Psychonomic Society, Inc. 2018

## Abstract

This paper presents a set of auditory perception experiments testing the effects of lexical ambiguity, lemma frequency, and acoustic details. In an AX discrimination task with stimuli produced in isolation, lexically ambiguous phonologically matching forms (e.g., *sun-sun*, *sun-son*) were evaluated more slowly and identified as ‘different’ more often than lexically unambiguous forms (e.g., *cat-cat*). For stimuli extracted from meaningful sentences, pairs of homophone mates (*sun-son*) were evaluated more slowly than same pairs of such words (*sun-sun*), following from the greater acoustic distance between homophone mates in several measures. The individual lexical frequency of homophone mates was a significant factor in both identification tasks, though frequency effects in the AX tasks were weaker and driven by the lexically unambiguous items. In both studies, greater acoustic distance between items was a predictor of longer response times, though the significance of particular acoustic measures varied. Identification of homophone mates also depended on context of production; listeners were above chance accuracy for choosing between homophone mates extracted from sentences, but not for homophone mates produced in isolation. While results for stimuli produced in sentential contexts indicate that listeners are sensitive to acoustic details and can weakly associate production patterns with lexical items, the absence of such differences for homophone mates produced in isolation suggests that these details are not an inherent part of the representation.

**Keywords** Phonology · Speech perception · Psycholinguistics

## Introduction

The semantic representations of homophones must be separate, given that they have different meanings (cf. Jastrzembski, 1981; Levelt, Roelofs, & Meyer, 1999), but there are two main types of models for the phonological representation of homophone mates. Either they share their phonological representation (Dell, 1990; Jescheniak & Levelt, 1994; Levelt et al., 1999), or they have separate but identical phonological representations (Gaskell & Marslen-Wilson, 1997; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982).

There is clear evidence for a representational level at which homophones are separate. Frequency effects on response latency are better predicted by the frequency of individual items than by homophone mates’ combined frequency in translation and picture naming (Caramazza, Costa, Miozzo, & Bi, 2001; Simpson & Krueger, 1991), lexical decision

tasks (Grainger, Van Kang, & Segui, 2001; Simpson & Burgess, 1985), and gaze duration in reading (Binder & Rayner, 1998). In addition, homophones have been found to prime themselves but not their homophone mates, under some conditions (Masson & Freedman, 1990; Schvaneveldt, Meyer, & Becker, 1976).

There is also evidence for connections between homophone mates. High-frequency words contribute to frequency effects of their low-frequency homophone mates in translation and picture naming (e.g., Antón-Méndez, Schütze, Champion, & Gollan, 2012; Jescheniak & Levelt, 1994) and lower susceptibility to production errors (Dell, 1990). The association can also be observed outside of frequency effects. Activation of one homophone can produce activation of the other, as is apparent in eye-tracking in reading tasks (Duffy, Morris, & Rayner, 1988) and in orthographic priming of lexical decisions (Onifer & Swinney, 1981).

Some work suggests that homophone mates have phonological representations that are not only separate but also distinct, based on the existence of observable phonetic differences in production (Gahl, 2008; Guion, 1995; Lohman, 2017). However, these phonetic differences might be due to influences in production. If they reflect differences in

✉ Chelsea Sanker  
chelsea\_sanker@brown.edu

<sup>1</sup> Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA

underlying representations, it should be possible to find evidence from perception that listeners associate these phonetic details with particular lexical entries.

### Accessing different levels of representation

Within models of representation, results might still vary due to different tasks accessing different levels of representation. The medium and context have effects on what patterns can be observed. In some contexts, both homophone mates are activated because there is no way to disambiguate between them, e.g., for homophones presented aurally without sentential context (Grainger et al., 2001; Onifer & Swinney, 1981). The results from orthographic tasks might not be paralleled in auditory experiments because of differences in ambiguity and corresponding activation.

Differences in the effect of lexical ambiguity depending on task type may suggest that homophones share phonological representations, though they have separate lexical representations. In many models, perceptual searches end when narrowed down to a single lexical entry (e.g., McQueen, Norris, & Cutler, 1994; Vitevitch & Luce, 1999). However, with acoustically presented homophones, the search cannot produce a single lexical item. In tasks that require semantic processing, putting homophone mates in competition, responses are slower for words with homophones (Hino, Lupker, & Pexman, 2002; Siakaluk, Pexman, Sears, & Owen, 2007) and can be slower further by priming their homophone mates (Pylkkänen, Llinás, & Murphy, 2006). In tasks in which the response is consistent for all homophone mates, e.g., lexical decision and naming, responses are faster for homophones than for other words (Borowsky & Masson, 1996; Hino et al., 2002; Kawamoto, Farrar, & Kello, 1994). The homophone advantage in the latter type of task can reflect an orthographic or phonological search strategy; consistent with this, the presence of phonotactically licit non-words, which reduce the viability of such a strategy for lexical decisions, eliminates the advantage for words with homophones (Borowsky & Masson, 1996; Davelaar, Coltheart, Besner, & Jonasson, 1978). The homophone advantage is sometimes interpreted as reflecting activation of multiple lexical entries with the same phonological form (Jastrzembski, 1981; Kellas, Ferraro, & Simpson, 1988), though this does not require that the phonological representations are also separate.

Frequency effects might also exhibit the two patterns of activation. If listeners are activating the lemma level, then frequency effects should follow homophone mates' individual frequencies. For acoustic stimuli, in which listeners cannot disambiguate between homophone mates, it is not clear which homophone's frequency will provide a better measure. Based on models of frequency effects on access, the higher frequency homophone should be accessed more

readily (cf. Binder & Rayner, 1998; Simpson & Burgess, 1985), so its frequency should be the one that produces the best model of word frequency. However, given evidence that ambiguous acoustic stimuli activate both homophone mates (Grainger et al., 2001; Onifer & Swinney, 1981), processing might continue until both homophones are activated, making the frequency of the lower frequency homophone a stronger predictor. If listeners are accessing only a shared level of phonological-level wordforms, frequency effects should follow homophone mates' joint frequency. However, lexical frequency effects are not always apparent in perception tasks that don't require semantic activation; some studies find an effect (e.g., Connine, Titone, & Wang, 1993; Howes, 1957), though others do not (e.g., Samuel, 1981).

The time course of effects also suggests both a shared stage of representation and an independent stage. Orthographically presented homophone mates prime each other at short delays (Lukatela & Turvey, 1994; Seidenberg et al., 1982; Tanenhaus, Leiman, & Seidenberg, 1979), but not at longer delays (Masson & Freedman, 1990; Schvaneveldt et al., 1976), suggesting an early stage of shared activation, followed by the suppression of one homophone mate. Lexical frequency effects occur later than effects of phonotactic probability, as is seen in MEG data (Pylkkänen, Stringfellow, & Marantz, 2002; Simon, Lewis, & Marantz, 2012), which suggests that lexical frequency effects primarily arise at this later stage, when the lemma level is activated.

### Interaction with orthography

Lexically ambiguous phonologically matching items do not necessarily form a uniform category. Multiple studies have found differences between homophone mates, with no semantic relationship, and the different forms of polysemous words, which are semantically related. On the other hand, studies on the effects of spelling on semantically unrelated homophone mates vary in whether or not they find a difference.

Many studies on homophones use orthographic stimuli, so results might reflect differences depending on whether listeners are accessing phonological or orthographic representations. Some studies have found that homographic homophones and heterographic homophones behave similarly, e.g., in response latencies in picture naming (Caramazza et al., 2001) and in aphasia patients' picture-naming accuracy after training (Biedermann & Nickels, 2008). However, other studies have found differences; gaze patterns during reading indicate more time spent disambiguating homographic homophones than heterographic homophones (Folk & Morris, 1995) and homographic homophone mates prime each other in picture naming, while heterographic homophones do not (Wheeldon & Monsell, 1992).

Homophones and polysemous words have different patterns in orthographic lexical decision tasks, suggesting that homophones have separate lemma representations, while polysemous forms are associated with the same lemma. Polysemous meanings tend to facilitate each other, producing faster responses, while homophones interfere with each other (Beretta, Fiorentino, & Poeppel, 2005; Klepousniotou & Baum, 2007; Rodd, Gaskell, & Marslen-Wilson, 2002). Neural activation also suggests shared lexical entries for meanings of polysemous words, while homophone mates have separate entries (Klepousniotou, Pike, Steinhauer, & Gracco, 2012; Pykkänen et al., 2006). Some of the apparent differences between heterographic and homographic homophones might in fact be cases of homophones vs. polysemously related forms.

### Acoustic details in representations

Homophones can exhibit significant phonetic differences in production due to word frequency (e.g., Gahl, 2008; Guion, 1995) and part of speech (e.g., Conwell, 2017; Sorensen, Cooper, & Paccia, 1978). If these differences are part of the representation, it would seem that some apparent homophone mates are actually cases of marginal phonological contrasts, phonetic differences which are less systematic than phonological contrasts either in perception or production, but nonetheless have reliably measurable differences (e.g., Renwick & Ladd, 2016; Scobbie & Stuart-Smith, 2008). Hall (2013) provides an overview of work that has identified and characterized such relationships; they are widely attested.

However, the differences can mostly be attributed to context, suggesting that they are not actually part of the underlying representation. Differences due to part of speech are largely attributable to phrase structure, as they are absent when items are placed in the same positions, e.g., sentence-finally (Conwell, 2017) and phrase-finally (Sorensen et al., 1978). The effects of lexical frequency have also sometimes been found to disappear when words are produced in isolation or in frame sentences (Guion, 1995); the most reliable effects are found within corpora of natural speech (e.g., Gahl, 2008; Lohman, 2017), and can be eliminated by controlling for other factors such as predictability based on context (Jurafsky, Bell, & Girand, 2002).

If phonetic differences between homophone mates are part of the underlying representation, the association should also be apparent in tests of perception. However, Bond (1973) found that listeners were at chance accuracy for identifying homophone mates. This result might have been influenced by aspects of the particular task, such as the lack of phonologically unambiguous filler items or the particular set of homophone mates, as there were only ten pairs and all pairs differed in morphological complexity.

Other perceptual tasks might find evidence of listeners attending to differences between homophone mates, if aspects of production result in consistent differences between them. Although listeners exhibit strongly categorical perceptual patterns, they are influenced by variation within categories. Reaction times are longer for identification and discrimination when the stimuli are closer to a category boundary and when paired items from different phonological categories are more similar (Pisoni & Tash, 1974); eye tracking exhibits similar effects (McMurray, Tanenhaus, & Aslin, 2002). Phonetic prototypicality also influences degree of lexical activation, as reflected in priming effects (Andruski, Blumstein, & Burton, 1994).

### Current study

The present study examines lexical representation and access using a set of auditory perception experiments. In particular, I look at effects of lexical ambiguity, lexical frequency, and phonetic detail, comparing homophones and other words. This new approach fits in with existing research on the representation of homophones based on speech production and responses to orthographic stimuli. Given that homophones are usually disambiguated in reading either by spelling or by context or both, perception of acoustically presented homophones might capture different patterns.

Experiments 1a and 2a were AX discrimination tasks, in which listeners heard pairs of items and decided whether they were ‘the same’ or ‘different’. This included pairs of the same word for words which have homophones (e.g., *sun-sun*) and don’t (e.g., *cat-cat*), and pairs of homophone mates (e.g., *sun-son*); there were an equal number of filler trials with phonologically distinct pairs. Experiments 1b and 2b were identification tasks, in which listeners identified an auditory stimulus as matching one of two written forms. The pairs were included homophone mates (e.g., *sun-son*) and phonologically distinct filler pairs. Experiments 1a and 1b used a set of stimuli extracted from meaningful sentences; Experiments 2a and 2b used a set of stimuli produced in isolation. Table 1 summarizes the predicted results for these tasks, depending on the representation of homophones and the level of representation being accessed.

Hypothesis 1 (lexical ambiguity): Effects of lexical ambiguity on responses can indicate what representation levels are activated by the task and shed light on possible representations, in particular whether the phonological representations of homophone mates are shared or if they are separate but identical.

- (a) If the semantic level is being activated in the auditory discrimination tasks, responses should be slower when two lemmas are activated (e.g., *sun-sun*, *sun-son*) than when a single lemma is activated (e.g., *cat-cat*),

**Table 1** Summary of primary hypotheses and predictions

|                                      | Shared Phonological Representations   |   | Separate Phonological Representations   |  |
|--------------------------------------|---|---|---|--|
| Acoustic differences in production   | No difference between <i>sun-sun</i> , <i>sun-son</i> , except perhaps due to sentential context            |   | Likely greater distance between <i>sun-son</i> than <i>sun-sun</i> , consistent across contexts |  |
|                                      | Phonological task strategy  | Lexical task strategy   | Phonological task strategy  | Lexical task strategy  |
| Lexical ambiguity (AX tasks)         | No effect   | Slower RT, fewer 'same' responses to <i>sun-sun</i> , <i>sun-son</i> than <i>cat-cat</i>                    | No effect   | Slower RT, fewer 'same' responses to <i>sun-sun</i> , <i>sun-son</i> than <i>cat-cat</i> |
| Homophone mates (AX tasks)           | Possible differences between <i>sun-sun</i> and <i>sun-son</i> , if acoustic differences exist              | Possible differences between <i>sun-sun</i> and <i>sun-son</i> , if acoustic differences exist              | Slower RT, fewer 'same' responses to <i>sun-son</i> than <i>sun-sun</i>                         | Slower RT, fewer 'same' responses to <i>sun-son</i> than <i>sun-sun</i>                  |
| Recording context (AX tasks)         | Slower RT, fewer 'same' responses to <i>sun-son</i> type possible only for stimuli extracted from sentences | Slower RT, fewer 'same' responses to <i>sun-son</i> type possible only for stimuli extracted from sentences | Slower RT, fewer 'same' responses to <i>sun-son</i> regardless of recording context             | Slower RT, fewer 'same' responses to <i>sun-son</i> regardless of recording context      |
| Frequency effects of homophone mates | Shared frequency  | Independent frequency   | Independent frequency   | Independent frequency  |
| Accuracy (identifications)           | Not feasible  | Possibly >50%, depending on recording context   | Not feasible  | >50%, consistent across recording contexts   |

regardless of whether the phonological representations of homophone mates are shared or independent. On the other hand, a lack of difference in response time would suggest that the task is approached at a lexically independent phonological level. It is also possible that lexically ambiguous items would be discriminated more rapidly than other items, as has been observed in other work with tasks that elicit phonological search strategies (e.g., Hino et al., 2002; Kawamoto et al., 1994), based on having multiple lexical entries contributing to their phonological activation.

- (b) The relationship between response time and lexical frequency can also be informative; if individual frequency of homophone mates is a better predictor of response time, this would suggest that listeners are activating independent representations, while if combined frequency of homophone mates is a better predictor, it would suggest that listeners are activating a shared representation.

Hypothesis 2 (differences between homophone mates):

- (a) If homophone mates are acoustically identical, response patterns in the discrimination tasks to pairs

of homophone mates (e.g., *sun-son*) and pairs with the same homophone (e.g., *sun-sun*) should be the same. If homophone mates have phonetic differences, but remain in the same phonological category, response times should be longer for pairs of homophone mates than for other pairs, given previous work demonstrating gradient effects of acoustic distance (e.g., McMurray et al., 2002; Pisoni & Tash, 1974), though the proportion of 'same' decisions is less likely to exhibit a difference, because the stimuli were all produced naturally and thus fall into the normal range of licit realizations of each category.

Greater acoustic distance that produces categorical distinctions should result in a larger number of 'different' responses to pairs of homophone mates than to other pairs. Perceptual differences between homophone mates and other pairs should also be paralleled by greater acoustic distance between items in these pairs.

- (b) Identification decisions with pairs of homophone mates in Experiments 1b and 2b provide another measure of whether or not they have contrastive differences; if there are no differences, or differences

are not salient, accuracy should be at chance (50%), and the only significant factors should be about selection preference (e.g., bias towards the item on the left), while differences which listeners associate independently with each homophone mate should produce substantially above-chance accuracy, which might interact with attentional factors.

**Hypothesis 3 (recording context):** The different contexts of production test how production environment influence the information available to listeners. If phonetic differences between homophone mates are present in the underlying representations, the context of elicitation should not matter; homophone mate pairs should exhibit slower responses and a larger number of “different” responses than other pairs in both discrimination tasks, and should be identified with above-chance accuracy in both identification tasks.

If phonetic differences between homophone mates arise due to production environments, different behavior of homophone mates and pairs of the same word should be apparent in Experiment 1a, with words extracted from meaningful sentences, but not in Experiment 2a, with words produced in isolation. If there are no phonetic differences between homophone mates in either context, pairs of homophone mates should not exhibit different results than other pairs in either discrimination task, and accuracy in identification tasks should never be above chance.

**Hypothesis 4 (filler type):** The three different block types based on the filler trials (differences in onsets, nuclei, or codas) test whether these fillers set up expectations about where in the words contrasts would appear and whether listeners were attending to particular subparts of homophone mates as loci for discriminating between them.

- (a) An influence of expectations set by the fillers would be reflected in differences in response times across block types, with the fastest responses in Onset blocks and the slowest responses in Coda blocks: In Onset blocks, listeners only need to hear the onset of the second word to determine whether or not the items match, while in Coda blocks, they cannot evaluate a match until they hear the full second word.
- (b) If differences across blocks indicate that fillers are directing listeners’ attention to particular parts of the word, an interaction with pair type could indicate that listeners are more uncertain about contrast boundaries of certain sounds or positions than about others. In particular, disproportionately slower responses to lexically ambiguous items in one type of block would suggest that listeners expect homophone contrasts to lie in that position.

## Study 1

### Experiment 1a: AX Discrimination Task

#### Methods

**Participants** 24 native speakers of American English (12 male; mean age 21) participated in this task and were paid for their participation.

**Materials** Stimuli were pairs of English words, extracted from meaningful sentences with similar phonological and syntactic environments. These productions were elicited orthographically, in randomized order, and said twice in succession; sentences were definitional, to produce similar contexts for each item (e.g., *to be fair is to be just, a fare is a fee; a doe is a deer, a dough is a mixture*). The target word was taken from the second utterance of each sentence, to ensure fluency and activation of the meaning of the target homophone. Each word in the pair came from a different speaker, from a set of three speakers. There were 46 words with homophones and 80 words that lack homophones. The list of homophones is given in the [Appendix](#).

The pair types were: (a) homophone-homophone (hph-hph) pairs (e.g., *sun-son*); (b) same pairs for a word with a homophone (e.g., *sun-sun*); (c) same pairs for a word with no homophone (e.g., *cat-cat*). Half of the homophones were homographic (e.g., *bank* ‘side of a river’ vs. *bank* ‘financial institution’) and half were heterographic (e.g., *sun* vs. *son*). For all pairs of words, the two items were produced by different speakers.

The frequency distribution of words without homophones was selected to approximate the frequency distribution of the heterographic homophones, for which there are more reliable frequency measures than are available for homographic homophones.

There were also filler trials of different pairs. There were three types of contrasts: onsets (e.g., *game-came*), nuclei (e.g., *look-lock*), and codas (e.g., *bed-bet*).

**Procedure** Pairs of words were presented over headphones, separated by 200 ms of silence. The experiment was run in PsychoPy (Pierce, 2007). Instructions on a screen asked listeners to decide whether the words were the same or different. Pilot testing demonstrated that listeners reliably interpreted this task as targeting phonological identity and not phonetic identity, given that all pairs differed in speaker and were thus phonetically distinct. Responses were given with the left and right arrow keys on the keyboard; which side corresponded to ‘same’ and ‘different’ was balanced across listeners. The next trial began 500 ms after a response was given. Response times were measured from the beginning of the second word.

Each block contained 160 pairs; the order of presentation was randomized. Within a block, contrasts in the fillers were always in the same position (onset, nucleus, or coda). All of the words that appeared in phonologically matching pairs also appeared in phonologically distinct filler pairs within the block. The phonological characteristics of the items in the lexically unambiguous pairs (type c) were balanced as much as possible with the characteristics of the lexically ambiguous pairs (types a-b). Each participant completed each of the three block types; the order in which they were presented was balanced across participants.

Half of participants heard no hph-hph pairs; the other half heard no same pairs for words with homophones (e.g., *sun-sun*). This restriction was based on pilot results indicating that hph-hph pairs were identified as being the same as frequently as same pairs were, and was aimed at balancing the number of phonologically matching pairs and phonologically distinct pairs.

All statistical results are from mixed effects models, calculated with the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). *p*-values were calculated by the lmerTest package (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2015). Responses with latencies shorter than 10 ms and the slowest 1% of responses (above 4.1 s) were excluded from analysis.

## Results

Accuracy was high, but not quite at ceiling. Pairs of homophone mates were identified as ‘same’ almost as consistently as pairs with the same word (93.0 vs. 93.9%). In comparison, the phonologically distinct filler pairs were identified as ‘same’ 6.2% of the time. Listeners made faster decisions in trials with the same pairs (*sun-sun*, *cat-cat*) than with different pairs (*sun-fun*, *cat-pat*): 1086 vs. 1201 ms.

Accurate responses were faster than inaccurate responses for same pairs (1061 vs. 1474 ms) and phonologically different pairs (1185 vs. 1444 ms). For hph-hph pairs, responses of ‘same’ were faster than responses of ‘different’ (1145 vs. 1518 ms). Based on the strong patterning of hph-hph pairs with the other pairs of phonologically matching items, these three pair types are considered in comparison to each other. All results reported in the following sections are only from these three pair types.

**Effects of block and pair type** Effects on listener decision patterns were small; the high rate of ‘same’ responses to all phonologically non-contrastive pairs left little room for variability due to other factors. Table 2 presents a generalized linear mixed effects model for ‘same’ responses to the phonologically matching stimuli. The random effects were participant and word pair (pairs of homophone mates

**Table 2** glmer model for ‘same’ responses in phonologically matching pairs, Exp. 1a

|                       | $\beta$ | SE    | z value | p value   |
|-----------------------|---------|-------|---------|-----------|
| (Intercept)           | 2.53    | 0.36  | 7.04    | <0.001*** |
| PairType Hph-Hph      | −0.081  | 0.20  | −0.41   | 0.68      |
| PairType Same non-hph | 0.47    | 0.24  | 1.91    | 0.056     |
| ContrastType C        | −0.076  | 0.14  | −0.55   | 0.58      |
| ContrastType O        | −0.21   | 0.14  | −1.51   | 0.13      |
| BlockNumb             | 0.14    | 0.067 | 2.10    | 0.035*    |
| WordFreq              | 0.027   | 0.055 | 0.49    | 0.62      |
| Homophone Type HG     | 0.079   | 0.26  | 0.31    | 0.76      |

Intercept: Type = Same (hph); ContrastType = N; HphType = non-homograph (nHG)

and corresponding same pairs of one of those homophones were treated as the same by this factor; i.e., *sun-sun* and *sun-son* have the same designation). The fixed effects were pair type (same hph, e.g., *sun-sun*; hph-hph, e.g., *sun-son*; non-hph, e.g., *cat-cat*); homophone type (homographic; non-homographic); contrast type of fillers within the block (Onsets; Nuclei; Codas); block number; and the log frequency of the higher frequency word in the pair.

The by-pair variance was 0.54 and the by-participant variance was 0.34.

There were slightly more ‘different’ responses to hph-hph pairs (7.0%) than to lexically unambiguous phonologically matching pairs (5.5%). The rate of ‘different’ responses to lexically ambiguous pairs of the same word was comparable to that of hph-hph pairs (7.2%). The difference cannot be explained as the result of differences in the mean acoustic distance, and instead is likely to reflect an effect of lexical ambiguity; when listeners’ evaluation of the auditory stimuli involves lexical access, lexically unambiguous pairs activate only one lemma, while the lexically ambiguous pairs activate two, which can increase the likelihood that the listeners will identify the items as being different.

There was no effect of homophone type (homographic or non-homographic).

There was no significant effect of contrast type on the portion of ‘same’ responses. That is, the position of phonological contrasts within the filler trials did not have an effect on the response pattern within the phonologically matching trials.

There was no effect of any measure of lexical frequency, nor any trend. In Table 2, the log frequency of the higher frequency homophone is used as the measure of frequency for words with homophones. Using the lower frequency homophone or the combined frequency of both homophone mates produced a similarly negligible result.

**Table 3** Lmer model for log response times in phonologically matching pairs, Exp. 1a

|                       | $\beta$ | SE     | t value | p value   |
|-----------------------|---------|--------|---------|-----------|
| (Intercept)           | 0.32    | 0.046  | 7.16    | <0.001*** |
| PairType Hph-Hph      | 0.066   | 0.023  | 2.93    | 0.0034**  |
| PairType Same non-hph | 0.0019  | 0.024  | 0.079   | 0.94      |
| ContrastType C        | 0.093   | 0.011  | 8.12    | <0.001*** |
| ContrastType O        | -0.011  | 0.012  | -0.92   | 0.36      |
| Response-Different    | 0.24    | 0.020  | 12.38   | <0.001*** |
| BlockNumb             | -0.060  | 0.0057 | -10.52  | <0.001*** |
| WordFreq              | -0.0032 | 0.0049 | -0.64   | 0.52      |
| Homophone Type HG     | 0.018   | 2.78   | 0.66    | 0.51      |
| Type Hph-Hph :        | -0.060  | 0.027  | -2.23   | 0.026*    |
| HphType HG            |         |        |         |           |

Intercept: Type = Same hph; ContrastType = N; Response = 'same'; HphType = non-homograph (nHG)

The proportion of 'same' responses to these phonologically matching pairs increased across blocks; this might suggest that familiarization with the speakers improved listeners' ability to identify phonological contrasts and how they map across the speakers.

There were stronger effects on response time; processing effects can be apparent from response latencies even when the final answers selected reflect only phonological categories. Table 3 summarizes effects on response time among phonologically matching pairs. The random effects were participant and word pair. The fixed effects were pair type (same hph, e.g., *sun-sun*; hph-hph, e.g., *sun-son*; non-hph, e.g., *cat-cat*); homophone type (homographic; non-homographic); contrast type of fillers within the block (Onsets; Nuclei; Codas); response ('same'; 'different'); block number; and the log frequency of the higher frequency word in the pair. The interaction between homophone type and pair type was also included.

The by-pair variance was 0.0057 and the by-participant variance was 0.019.

Response was a significant predictor of response time, as discussed above: decisions of 'different' for phonologically matching items were slower than decisions of 'same'. This pattern was present for all pair types in the model. Response was included as a factor, rather than being used to exclude inaccurate responses, in order to avoid building the model based on any assumptions about whether pairs of homophone mates should accurately be identified as being the same or different.

Responses to hph-hph pairs were significantly slower than responses to same pairs for words with homophones (1172 vs. 1085 ms). This effect must be due to differences

in production, which will be summarized in Table 4. In all pair types, the two items were produced by different speakers, so the comparison was always between different voices. However, hph-hph pairs had slight differences in their production environments, while the tokens for same pairs came from identical sentences.

Among pairs of the same word, there was no significant difference between words with homophones (e.g., *sun-sun*; mean RT = 1085 ms) and those without (e.g., *cat-cat*; mean RT = 1087 ms). The lack of difference between these types further indicates that the slow response times for the hph-hph pairs were due to acoustic details, not lexical ambiguity. Figure 1 presents the response time distributions, by pair type.

Among lexically ambiguous pairs, there was no significant effect of whether homophone mates had different spelling (e.g., *sun-son*) or the same spelling (e.g., *bank* 'side of a river' - *bank* 'financial institution'). However, there was a weak interaction with pair type. In hph-hph pairs, responses were slightly slower to heterographic homophones; in same pairs, responses were slightly faster to heterographic homophones. The difference may suggest that acoustic variation is more likely to be interpreted as potentially contrastive in phonological forms associated with multiple distinct orthographic forms. However, the difference is small relative to the variability in response times within each category.

Even though different pairs were excluded from calculations, response times differed depending on where the differences were in the phonologically distinct filler pairs within the block, directing listeners' attention at different parts of the words. Based on expectations set up by the fillers, listeners could make decisions sooner when differences could be expected to occur earlier in the word. Responses were fastest for onset decisions (1066 ms) and nucleus decisions (1073 ms), with slower responses for coda decisions (1184 ms); only the response times for blocks with coda contrasts differed significantly from other blocks. This is consistent with most onsets and nuclei becoming clear during the transition into the vowel, while the coda is not identifiable until the end.

If listeners are more likely to look for contrasts between homophones as residing in a particular part of the word, or find low-level acoustic differences more difficult to resolve in some positions than in others, pair type might interact with block type. However, an interaction between pair type and block-level contrast type did not improve the model ( $\chi^2 = 1.67$ ,  $df = 4$ ,  $p = 0.80$ ).

Lexical frequency was not significant in the model. Word frequencies are based on counts from the part-of-speech tagged Brown Corpus (Francis & Kučera, 1982). While this corpus is smaller than other corpora that provide word

**Table 4** Mean acoustic distance between paired items in phonologically matching pairs, Exp. 1a. Standard deviations are given in parentheses

|              | Euclidian dist.  | Vowel dur. diff. | F0 max diff.   | spec. tilt diff. |
|--------------|------------------|------------------|----------------|------------------|
| Hph-Hph      | 265.6 Hz (143.4) | 41.5 ms (24.1)   | 62.3 Hz (32.9) | 5.2 dB (4.6)     |
| Same Hph     | 226.9 Hz (117.3) | 39.9 ms (30.4)   | 54.5 Hz (36.3) | 4.9 dB (4.2)     |
| Same Non-Hph | 204.4 Hz (115.9) | 40.9 ms (31.8)   | 56.9 Hz (37.3) | 4.9 dB (4.2)     |

frequency information, the part of speech tagging made it possible to distinguish between homographic homophones. Frequency counts for homographic homophone mate pairs which are not distinguished by part of speech were made by manually categorizing each instance of the form. The log frequency of the higher frequency homophone was used for words with homophones in the model presented in Table 3. Using the lower frequency homophone or the combined frequency of both homophone mates produced a similar non-significant result. The null result may be because the data set was not sufficiently large or did not cover a broad enough range of frequencies, though it is also possible that frequency effects depend on the particular task and that this task simply was not affected by lexical frequency.

Response times decreased across blocks, which is consistent with listeners acclimating to the task and being able to respond more quickly. As seen in Table 2, the proportion of ‘same’ responses to these pairs also increased, suggesting an overall increase in proficiency in the task.

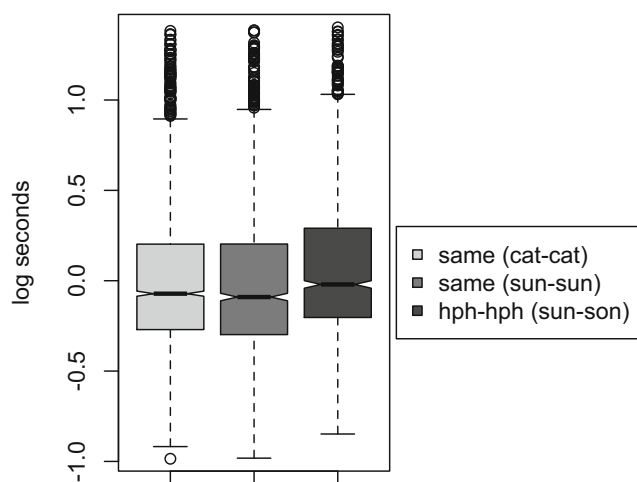
**Effects of acoustic details** Acoustic measurements were taken for each pair using Praat (Boersma & Weenink, 2017). While there were no significant differences in the mean acoustic difference between paired items that were elicited as the same word from two speakers vs. homophone mates from two speakers in any of the measures used (F1 x F2 distance, vowel duration, maximum F0, spectral tilt), there were trends towards greater acoustic distance between

homophone mates; a summary is presented in Table 4. Formant measures included only monophthongs. Spectral tilt was measured as H1-H2, taken in the middle of the vowel.

Greater differences between homophone mates can result from differences in word frequency (Gahl, 2008; Guion, 1995) and part of speech (Lohman, 2017; Sorensen et al., 1978). All target words occurred in the same phrasal environment, as the last word of the subject noun phrase or infinitive that was followed by a copula and definitional predicate (e.g., *the sun is a star*; *to meet is to encounter*). Though the exact sentence was different for homophone mates, the overall similarity in position and sentence length should have prevented large differences in their form. Recall that all pairs had different speakers for the two items, so the task was never to identify matching tokens or matching voices. Much of the variation in distance between items is driven by differences across speakers, though listeners’ high accuracy of identifications indicated that they were able to reliably account for these differences.

Listeners are likely attending to a range of characteristics, and might be influenced by additional characteristics not included here. Nonetheless, the pattern of greater distance between the paired items in hph-hph pairs suggests that the differences in perceptual behavior resulted from acoustic differences. Words without homophones patterned much like words with homophones, though there was a smaller distance in the vowel space of paired items in the former category. This difference might result from variation in the phonological environment within the word; while the set of hph-hph items had exactly the same segmental sequences as the set of items in which one of those homophone mates was paired with itself, the set of same pairs for words without homophones did not have this exact parallel.

The model for acoustic effects on response patterns illustrated in Table 5. The random effects were participant and word pair. The fixed effects were: pair type (same hph, e.g., *sun-sun*; hph-hph, e.g., *sun-son*; non-hph, e.g., *cat-cat*); contrast type of fillers within the block (Onsets; Nuclei; Codas); response (‘same’; ‘different’); block number; log frequency of the higher frequency word; and Euclidean distance between the vowels of the paired words. Homophone type was not included, because it did not improve the model and it obscured some of the other effects.

**Fig. 1** Response time by pair type, Experiment 1a



**Table 5** glmer model for ‘same’ responses in phonologically matching pairs, Exp. 1a

|                       | $\beta$ | SE      | z value | p value   |
|-----------------------|---------|---------|---------|-----------|
| (Intercept)           | 3.12    | 0.41    | 7.55    | <0.001*** |
| PairType Hph-Hph      | -0.27   | 0.22    | -1.20   | 0.23      |
| PairType Same non-hph | 0.33    | 0.24    | 1.38    | 0.17      |
| ContrastType C        | -0.033  | 0.15    | -0.22   | 0.83      |
| ContrastType O        | -0.18   | 0.15    | -1.23   | 0.22      |
| BlockNumb             | 0.15    | 0.074   | 2.06    | 0.039*    |
| WordFreq              | -0.029  | 0.061   | -0.48   | 0.63      |
| EuclidDist            | -0.0012 | 0.00069 | -1.79   | 0.073     |

Intercept: Type = Same (hph); ContrastType = N

The by-pair variance was 0.60 and the by-participant variance was 0.37.

Listeners gave fewer ‘same’ responses to pairs of items that differed more in their vowels’ formant structure, as measured in Euclidean distance in the F1 x F2 space. Because of the differences in acoustic distance across pair types, including both factors reduced the effects of each predictor; in a model without pair type, Euclidean distance was significant ( $\beta = -0.0017$ , z-value =  $-2.46$ ,  $p = 0.014$ ), rather than just marginal. None of the other acoustic measures approached significance, so they were not included in the model.

Table 6 presents a linear mixed effects model revised to include measures of acoustic distance. The random effects were participant and word pair. The fixed effects were: pair type (same hph, e.g., *sun-sun*; hph-hph, e.g., *sun-son*; non-hph, e.g., *cat-cat*); contrast type of fillers within the block (Onsets; Nuclei; Codas); response (‘same’; ‘different’); block number; log frequency of the higher frequency word;

**Table 6** lmer model for log response times in phonologically matching pairs, Exp. 1a

|                       | $\beta$  | SE       | t value | p value   |
|-----------------------|----------|----------|---------|-----------|
| (Intercept)           | 0.27     | 0.049    | 5.49    | <0.001*** |
| PairType Hph-Hph      | 0.043    | 0.021    | 2.10    | 0.036*    |
| PairType Same non-hph | -0.0035  | 0.021    | -0.17   | 0.87      |
| ContrastType C        | 0.095    | 0.013    | 7.49    | <0.001*** |
| ContrastType O        | -0.0079  | 0.013    | -0.61   | 0.54      |
| Response-Different    | 0.24     | 2.15     | 11.08   | <0.001*** |
| BlockNumb             | -0.064   | 0.0063   | -10.17  | <0.001*** |
| WordFreq              | 0.0017   | 0.0053   | 0.32    | 0.75      |
| EuclidDist            | 0.000098 | 0.000061 | 1.62    | 0.11      |
| VowelDurDiff          | 0.85     | 0.28     | 3.08    | 0.0025**  |

Intercept: ContrastType = N; Response = ‘same’

Euclidean distance between the vowels of the paired words; and difference in vowel duration of the paired words.

The by-pair variance was 0.0056 and the by-participant variance was 0.019.

Responses were slower when the items were more acoustically distinct in their formant structure or their vowel duration. That is, it took longer for listeners to decide whether phonologically matching items were the same when they were more acoustically different. The differences in acoustic distance across pair types meant that including both factors reduced the effects of each predictor; in a model without pair type, Euclidean distance was significant ( $\beta = 0.00013$ ,  $t$  value =  $2.09$ ,  $p = 0.037$ ). The significant effects of vowel duration and Euclidian distance between the formant structures suggest that listeners are attending to both of these characteristics as potential cues to phonological contrasts.

The main results for Experiment 1a are summarized in Table 7.

### Experiment 1b: Identification task

Given that there are acoustic differences between homophone mates, and listeners can be sensitive to them when forms are juxtaposed with each other, can listeners use these differences to identify homophones in isolation? If acoustic details are part of the representation of particular items, this association should be apparent in word identification tasks. Experiment 1b was aimed at examining these possibilities for the same items from Experiment 1a.

### Methods

**Participants** 24 native speakers of American English (nine male; mean age 21.3) participated in this task and were paid for their participation. All participants also

**Table 7** Summary of main results, Exp. 1a

|                           | ‘same’ responses              | Response times                 |
|---------------------------|-------------------------------|--------------------------------|
| Same hph vs. Hph-hph      | not sig.                      | slower for hph-hph             |
| Same hph vs. Same non-hph | higher for non-hph            | not sig.                       |
| Contrast type N vs. C     | not sig.                      | slower for C                   |
| Contrast type N vs. O     | not sig.                      | not sig.                       |
| Frequency                 | not sig.                      | not sig.                       |
| Response                  | NA                            | slower for ‘different’         |
| Euclidean dist.           | lower with greater difference | slower with greater difference |
| Vowel duration diff.      | not sig.                      | slower with greater difference |

participated in an AX discrimination task (Experiment 1a or Experiment 2a), which they completed prior to this task.

**Materials** Auditory stimuli were individual English words from Experiment 1a. They were associated with written options that were pairs from Experiment 1a, of the types: (a) homophone-homophone (hph-hph) pairs (e.g., *sun-son*) and two types of fillers, (b) different pairs in which one of the words has a homophone (e.g., *sun-fun*), and (c) different pairs in which neither word has a homophone (e.g., *cat-pat*). There were 20 pairs of homophone mates, all orthographically unambiguous. These are given in the Appendix, in the heterographic column for Study 1, excluding the three items that also have homographs. Only the results for homophone mates are analyzed, aside from a brief summary of filler results, which confirmed that participants were attending to the task.

**Procedure** The experiment was run in PsychoPy. Participants heard individual words played in isolation and identified each by selecting one of two written options. The written options appeared 500 ms before the auditory stimulus, which matched one of them. The early presentation of the written forms was based on pilot results suggesting that simultaneous presentation would facilitate a strategy in which listeners did not attend to both written forms, but instead selected the first written item they saw that was consistent with the auditory stimulus, which would not test their ability to decide between homophone mates. Early presentation of written forms reduced the left-side bias of responses, suggesting greater attention to both items. Across participants, the presentation positions of the two items were balanced, to control for possible preference either in use of the arrow keys or in attention to a screen side.

The next trial began 500 ms after a response was given. Response times were measured from the beginning of the word.

Both words from each pair appeared among the stimuli. The side of the screen where each word appeared was consistent for each participant (e.g., if *sun* and *son* appeared on the left and the right respectively when listeners heard the *sun* stimulus, they were also in that order for the *son* stimulus). The presentation positions of the two items were balanced across participants. The order of stimuli was randomized.

There were two conditions. For half of the participants, homophone stimuli were only presented for discrimination with their homophone mates (e.g., auditory *sun* with the written options *sun* and *son*), i.e., excluding filler type (b). In this condition, there were 120 trials in each block, with 40 trials deciding between homophone mates. For the other half of participants, homophones were presented both for discrimination from their homophone mates and from

phonologically distinct words (e.g., auditory *sun* with the written options *sun* and *fun*), i.e., including both filler types. In this condition, there were 200 trials in each block, with 40 trials deciding between homophone mates.

Each participant completed three blocks; blocks differed in where the difference in the phonologically distinct pairs lay: onsets (e.g., *game*, *came*), nuclei (e.g., *look*, *lock*), and codas (e.g., *bed*, *bet*). These pairs served both to ensure that listeners were not guessing randomly and also to direct listeners' attention to different parts of the words.

All statistical results are from a mixed effects model, calculated with the lme4 package in R (Bates et al., 2015). *p*-values were calculated by the lmerTest package (Kuznetsova et al., 2015). Responses with latencies shorter than 10 ms and the slowest 1% of responses in each category (>5.4 s for homophone mate decisions) were excluded from analysis.

## Results

Listeners' accuracy for decisions between phonologically distinct words was 97.4%, demonstrating that they understood the task and were paying attention. Response times were much slower for decisions about homophone mates (1718 ms) than about phonologically distinct words (1401 ms). The consistently longer response times for decisions about homophone mates suggest that listeners were actually considering these pairs, rather than simply guessing. Only homophone mate pairs are included in all subsequent analyses.

A generalized linear mixed effects model for accuracy of decisions between homophone mates is presented in Table 8. The random effects were participant and word pair. The fixed effects were: side of the screen on which the correct answer appeared (left; right); whether or not the

**Table 8** glmer model for accuracy in homophone identification

|                              | $\beta$ | SE      | z value | <i>p</i> value |
|------------------------------|---------|---------|---------|----------------|
| (Intercept)                  | 0.45    | 0.21    | 2.17    | 0.030*         |
| ScreenSide right             | −0.35   | 0.076   | −4.63   | <0.001***      |
| StimSet no hph-nonhph trials | −0.029  | 0.084   | −0.35   | 0.73           |
| ContrastType C               | −0.27   | 0.093   | −2.92   | 0.0036**       |
| ContrastType O               | −0.049  | 0.093   | −0.53   | 0.60           |
| ResponseTime                 | −0.023  | 0.052   | −0.46   | 0.65           |
| BlockNum                     | −0.036  | 0.047   | −0.76   | 0.45           |
| Trial                        | −0.0019 | 0.00082 | −2.33   | 0.020*         |
| FreqCorr                     | 0.090   | 0.021   | 4.37    | <0.001***      |
| FreqIncorr                   | −0.049  | 0.020   | −2.38   | 0.017*         |

Intercept: ScreenSide = left; StimSet = with hph-nonhph trials; ContrastType = N

trials for pairs of phonologically distinct words included words with homophones (with hph-nonhph trials, e.g., *sun-fun*; no hph-nonhph trials); contrast type of fillers within the block (Onsets; Nuclei; Codas); response time; block number; trial number within the block; log frequency of the acoustically presented item; and log frequency of the competing homophone mate.

The by-pair variance and by-participant variance were both negligible ( $<0.001$ ).

Taking other factors into account, accuracy was significantly above chance, though the overall accuracy was only 50.8%. While the size of the effect is small, this suggests that the acoustic details that produced effects in the discrimination task of Experiment 1a are not just salient in juxtaposition, but can be weakly associated with particular words.

The strongest effect was a preference for selecting the written option on the left side of the screen, so accuracy was higher when the correct answer appeared on the left. This result is likely based on English listeners reading from left to right, so the word on the left was more salient, even though the visual items were presented before the auditory stimulus to encourage awareness of both items. Stimuli were balanced to have an equal number of correct items appearing on the left and the right side of the screen, so this bias did not influence measurements of perceptually motivated accuracy.

There was no significant effect of whether homophones were only presented for discrimination with their homophone mates or also appeared with phonologically distinct items, that is, whether the *sun-fun* type fillers were present or not. This suggests that there was no training effect of having homophones appear in contexts which made clear which homophone mate was meant.

The position of the contrasts present in the phonologically contrastive filler pairs within the block also had a significant effect; accuracy within the homophone mate pairs was higher when other pairs within the block contrasted in nuclei or onsets, and lower when the contrast was in codas. This might suggest that attention to codas distracted listeners from the acoustic differences that could actually provide cues to discriminate between the homophone mates, while nuclei provided cues that actually align with differences between the homophone mates.

However, the effect of filler type was not due to differences in response time in each block. There was no effect of response time on accuracy; while response times were variable, with some decisions taking several seconds, longer deliberation was neither beneficial nor detrimental.

There was no significant effect of block number on accuracy. As the same homophone tokens were presented in each block, this indicates that exposure to these tokens was not beneficial in establishing accurate associations between

the particular forms and the words represented. The lack of effect across blocks of this task suggests that participation in a prior discrimination task with these homophones also did not have an effect of priming homophone contrasts or training listeners in particular items. Experiment 2b compares discrimination results for participants who did or did not first complete the discrimination task, to further test possible interaction between the two tasks.

Trial number was a significant predictor of accuracy; accuracy decreased within a block. This might suggest a fatigue effect, if listeners are capable of using word-specific details but have difficulty sustaining that level of attention for many trials, though notably the lack of effect of block number indicates that the decrease in accuracy across trials does not persist across blocks. The decrease in accuracy does not plateau at 50%, which might suggest that listeners are actually developing actively counter-productive strategies.

There was a significant effect of lexical frequency on decisions; listeners more often selected higher-frequency items as responses. This was apparent both as a positive effect of the frequency of the stimulus item and a negative effect of the frequency of the homophone mate. That is, the response matching the stimulus was more often selected when it was higher frequency, and when the stimulus word's homophone mate was higher frequency, that competitor decreased the chances of correct responses.

## Study 2

### Experiment 2a: AX Discrimination Task

#### Methods

**Participants** 24 native speakers of American English (eight male; mean age 22.1) participated in this task and were paid for their participation.

**Materials** Stimuli were pairs of English words, produced in isolation in response to orthographic stimuli in randomized order. Each word in the pair came from a different speaker, with two total speakers. There were 80 words with homophones, and 172 words without homophones. The homophones are given in the [Appendix](#).

The pair types were: (a) homophone-homophone (hph-hph) pairs (e.g., *sun-son*); (b) same pairs for a word with a homophone (e.g., *sun-sun*); (c) same pairs for a word with no homophone (e.g., *cat-cat*). All homophone pairs were non-homographic. For all pairs of words, the two items were produced by different speakers. Words were selected to capture a range of lexical frequencies, with similar frequencies in each pair type.

There were also filler trials of different pairs, each of which included at least one word which also appeared among the phonologically matching stimuli. There were three types of contrasts in the different pairs: onsets (e.g., *game-came*), nuclei (e.g., *look-lock*), and codas (e.g., *bed-bet*).

**Procedure** Presentation of stimuli and collection of responses followed the same procedure as Experiment 1a. Block design was also the same. The only difference was that all participants heard all pair types, for a total of 320 pairs in each block. As before, the number of phonologically matching pairs and phonologically distinct pairs was balanced.

All statistical results are from mixed effects models, calculated with the lme4 package in R (Bates et al., 2015). *p* values were calculated by the lmerTest package (Kuznetsova et al., 2015). Responses with latencies shorter than 10 ms and the slowest 1% of responses (above 4.9 s) were excluded from analysis.

## Results

As in Experiment 1a, hph-hph pairs patterned like same pairs, both in primarily eliciting ‘same’ responses and also in the timing of responses. They were regularly identified as ‘same’ (88.6% of responses), with a similar frequency as pairs of the same word (89.5%); in comparison, phonologically distinct words were identified as ‘same’ 4.2% of the time. Response times to phonologically different pairs were slightly faster (1155 ms) than to hph-hph pairs (1189 ms) or same pairs (1177 ms). There were also more ‘different’ responses than ‘same’ responses.

Accurate responses were significantly faster than inaccurate responses for same pairs (1129 vs. 1589 ms) and phonologically different pairs (1146 vs. 1357 ms); for hph-hph pairs, ‘same’ responses were faster than ‘different’ responses (1119 vs. 1726 ms), further establishing that they were perceived as being the same. Accuracy among phonologically different pairs was slightly higher in this experiment than in Experiment 1a. This likely reflected the stimuli being produced in isolation, so the context of production matched the listening context and reduced phonological ambiguity. However, accuracy in same pairs was lower, as will be discussed below.

Based on the strong patterning of hph-hph pairs with the other phonologically matching pairs, these three pair types are considered in comparison to each other. All results reported subsequently are just from these three pair types, as in Experiment 1a.

**Effects of block and pair type** There were more responses of ‘different’ to phonologically matching pairs than there were

**Table 9** glmer model for ‘same’ responses in phonologically matching pairs, Exp. 2a

|                   | $\beta$ | SE    | z value | p value   |
|-------------------|---------|-------|---------|-----------|
| (Intercept)       | 1.88    | 0.41  | 4.57    | <0.001*** |
| Type Hph-Hph      | 0.14    | 0.091 | 1.56    | 0.12      |
| Type Same Non-hph | 0.53    | 0.13  | 3.99    | <0.001*** |
| ContrastType C    | 0.021   | 0.086 | 0.25    | 0.81      |
| ContrastType O    | 0.081   | 0.088 | 0.92    | 0.36      |
| BlockNum          | -0.17   | 0.042 | -4.08   | <0.001*** |
| WordFreq          | 0.076   | 0.031 | 2.47    | 0.014*    |

Intercept: Type = Same Hph; ContrastType = N

in Experiment 1a, so some factors influencing response pattern appeared with greater strength. Table 9 presents a generalized linear mixed effects model for these effects. The random effects were participant and word pair. The fixed effects were pair type (same hph, e.g., *sun-sun*; hph-hph, e.g., *sun-son*; non-hph, e.g., *cat-cat*); contrast type of fillers within the block (Onsets; Nuclei; Codas); block number; and log frequency of the higher frequency item in the pair.

The by-pair variance was 0.29 and the by-participant variance was 1.11.

Same pairs for words without homophones were identified as ‘same’ significantly more frequently (90.6%) than hph-hph pairs (88.6%) or same pairs for words with homophones (87.3%); the difference between the latter two types was not significant. This significantly higher portion of ‘same’ responses among lexically unambiguous pairs than among other pairs suggests that listeners were using different decision strategies in this experiment than they were in Experiment 1a, perhaps more reliably activating

**Table 10** lmer model for log response times in phonologically matching pairs, Exp. 2a

|                          | $\beta$ | SE     | t value | p value   |
|--------------------------|---------|--------|---------|-----------|
| (Intercept)              | 0.36    | 0.062  | 5.88    | <0.001*** |
| Type Hph-Hph             | -0.039  | 0.031  | -1.27   | 0.20      |
| Type Same Non-hph        | -0.11   | 0.030  | -3.71   | <0.001*** |
| ContrastType C           | 0.046   | 0.0098 | 4.75    | <0.001*** |
| ContrastType O           | -0.010  | 0.0097 | -1.04   | 0.30      |
| Response-Different       | 0.20    | 0.024  | 8.47    | <0.001*** |
| BlockNum                 | -0.047  | 0.0046 | -10.12  | <0.001*** |
| WordFreq                 | -0.0059 | 0.0033 | -1.78   | 0.077     |
| TypeHph-Hph:ResponseSame | 0.045   | 0.033  | 1.37    | 0.17      |
| TypeNon-hom:ResponseSame | 0.12    | 0.030  | 4.03    | <0.001*** |

Intercept: Type = Same Hph; ContrastType = N; Response = ‘same’

lexical representations and not just phonological representations. Notably, the response times were also longer.

There was also a significant decrease in the portion of ‘same’ responses to the phonologically matching stimuli across the blocks, which might suggest a fatigue effect or suggest that listeners were developing strategies to try to differentiate between homophone mates, although there were no characteristics present which actually distinguish them. Note that this is in contrast to Experiment 1a, in which accuracy increased across blocks. This difference is likely to reflect the differences between the stimuli of the two tasks; in Experiment 1a, with stimuli extracted from sentences, there were acoustic differences that aligned with the different homophone mates, whereas in Experiment 2a, there is no evidence for greater acoustic differences between homophone mates than between items in other pairs, as will be discussed in the following section.

Lexical frequency was a predictor of accuracy, with higher accuracy for pairs containing higher frequency items. This effect is entirely driven by the lexically unambiguous pairs, and is absent in a model without those pairs ( $\beta = -0.035$ ,  $z$ -value =  $-0.74$ ,  $p = 0.46$ ).

Table 10 summarizes effects on response time among phonologically matching pairs. The random effects were participant and word pair. The fixed effects were pair type (same hph, e.g., *sun-sun*; hph-hph, e.g., *sun-son*; non-hph, e.g., *cat-cat*); contrast type of fillers within the block (Onsets; Nuclei; Codas); response (‘same’; ‘different’); block number; and log frequency of the higher frequency item in the pair. The interaction between response and pair type was also included. Most effects paralleled the results in Experiment 1a, with the notable exception of pair type, which exhibited a different split, in addition to a stronger effect of lexical frequency.

The by-pair variance was 0.0033 and the by-participant variance was 0.048.

Responses of ‘different’ were slower than responses of ‘same’. This interacted significantly with type. The higher number of ‘different’ responses in this experiment, likely reflecting a different response strategy, provided more potential for interactions with other factors to become apparent. Among same pairs for words without homophones, the difference in response time between responses of ‘same’ and ‘different’ was much smaller than

it was among same pairs for words with homophones or among hph-hph pairs.

Responses to same pairs for words without homophone mates were faster (e.g., *cat-cat*; mean RT = 1171 ms) than hph-hph pairs (e.g., *sun-son*, mean RT = 1189 ms) or same pairs for words with homophones (e.g., *sun-sun*; mean RT = 1191 ms); the difference between the latter two types was not significant. This pattern suggests that the response latency was reflecting an influence lexical knowledge, rather than any acoustic differences between homophone mates, in contrast to Experiment 1a, in which pairs of homophone mates were slower than the other two types of phonologically matching pairs. Figure 2 presents the response time distributions, by pair type and response; note that the difference in response time between lexically ambiguous items and unambiguous items was mostly apparent within the ‘different’ responses.

Consistent with the results from Experiment 1a, response times differed depending on which part of the word the block set up as being relevant for contrasts; responses were fastest for onset decisions (1162 ms), then nucleus decisions (1174 ms), then coda decisions (1204 ms). Again, the only significant difference was between blocks with coda differences and other blocks.

Including an interaction between contrast type and pair type did not improve the model ( $\chi^2 = 2.28$ ,  $df = 4$ ,  $p = 0.68$ ). This result suggests that even though lexical ambiguity slowed down responses and resulted in a larger number of ‘different’ responses, which could suggest a degree of uncertainty about whether or not there is an associated phonological contrast, listeners do not have consistent expectations about where within these items such contrasts might lie.

Word frequencies are based on counts in the Corpus of Contemporary American English (Davies, 2008); the set of stimuli was designed to contain a range of frequencies, with comparable distributions among homophones and non-homophones. This larger corpus was used to better capture differences among low frequency words, which is limited within the Brown Corpus used for frequencies in Experiment 1a; all homophones in Experiment 2a were heterographic, which obviated the need for the part-of-speech tagging provided by the Brown Corpus.

**Table 11** Acoustic distance between paired items in phonologically matching pairs, Exp. 2a

|              | Euclidian dist.  | Vowel dur. diff. | F0 max diff.   | spec. tilt diff. |
|--------------|------------------|------------------|----------------|------------------|
| Hph-Hph      | 199.9 Hz (153.0) | 55.1 ms (35.8)   | 73.6 Hz (22.7) | 8.7 dB (8.1)     |
| Same Hph     | 183.4 Hz (131.9) | 57.5 ms (32.4)   | 73.7 Hz (22.7) | 8.4 dB (7.9)     |
| Same Non-Hph | 181.9 Hz (125.6) | 56.0 ms (32.5)   | 73.9 Hz (27.2) | 8.6 dB (7.4)     |

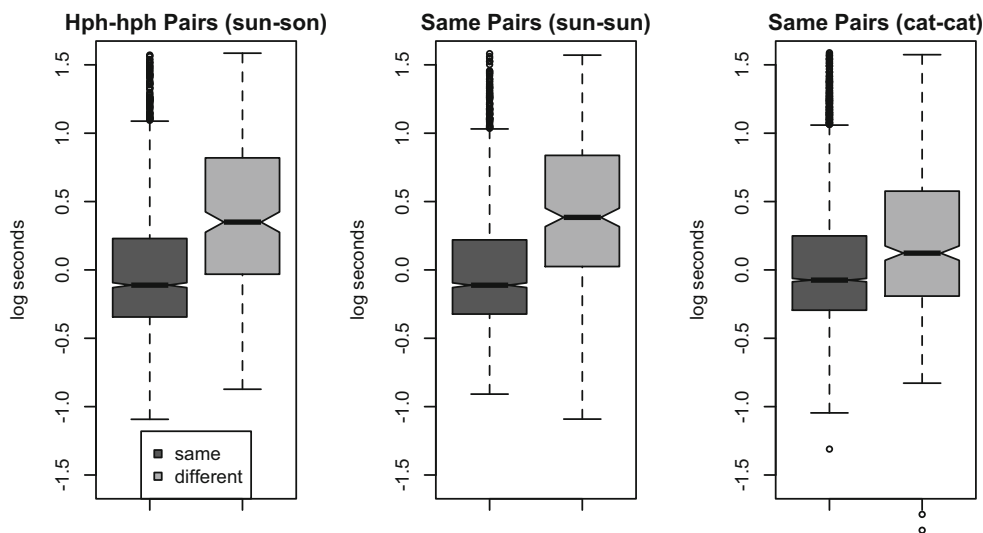


Fig. 2 Response time by pair type and response, Experiment 2a

Responses to higher frequency words were somewhat faster. This effect is driven by the lexically unambiguous pairs, and is absent in a model that excludes those pairs ( $\beta = 2.13$ ,  $z$ -value = 0.33,  $p = 0.74$ ).

Response time decreased across blocks, which is consistent with listeners acclimating to the task, though the improvement in speed notably was not accompanied by an improvement in response accuracy, in contrast to the results in Experiment 1a.

**Effects of acoustic details** The two items in each pair were not significantly more distinct in hph-hph pairs than in same pairs in any of the measures used (F1 x F2 distance, vowel duration, maximum F0, spectral tilt). Unlike in Experiment 1a, there was no tendency for larger acoustic differences between homophone mates than between instances of the same word, though the Euclidian distance in the vowel space was slightly greater. These measurements are summarized in Table 11.

There were no significant effects of acoustic distance on response patterns in this experiment, so no regression model is included to illustrate this absence of effects. Although listeners were influenced by lexical ambiguity and gave a larger number of ‘different’ responses in this experiment than in Experiment 1a, this lack of effect indicates that these decisions were not strongly influenced by the aspects of phonetic form investigated here.

Nonetheless, acoustic differences did predict response time. Table 12 presents the linear mixed effects model for response times, revised to include measures of acoustic distance. The random effects were participant and word pair. The fixed effects were pair type (same hph, e.g., *sun-sun*; hph-hph, e.g., *sun-sun*; non-hph, e.g., *cat-cat*); contrast type of fillers within the block (Onsets; Nuclei; Codas); response

(‘same’; ‘different’); log frequency of the higher frequency item; block number; difference in vowel duration between paired items; and difference in spectral tilt.

The by-pair variance was 0.0032 and the by-participant variance was 0.048.

Responses were slower when the items differed more in spectral tilt or in vowel duration. There was substantial overlap in the effects of each acoustic factor; the effects of F0 maximum and Euclidian distance, which approached significance in individual models, were reduced when other acoustic factors were included. As they did not improve the model, they were excluded from the model. Although the acoustic differences across pair types were smaller than in Experiment 1a, including both pair type and acoustic measures reduced some of the effects of each; in a model without pair type, the difference in vowel duration was a significant predictor ( $\beta = 0.49$ ,  $t$  value = 2.31,  $p = 0.021$ ).

Table 12 lmer model for log response times in phonologically matching pairs, Exp. 2a

|                    | $\beta$  | SE      | $t$ value | $p$ value  |
|--------------------|----------|---------|-----------|------------|
| (Intercept)        | 0.27     | 0.061   | 4.47      | < 0.001*** |
| Type Hph-Hph       | -0.00021 | 0.011   | -0.020    | 0.98       |
| Type same Non-hph  | -0.0047  | 0.014   | -0.33     | 0.74       |
| ContrastType C     | 0.046    | 0.0097  | 4.70      | < 0.001*** |
| ContrastType O     | -0.010   | 0.0097  | -1.05     | 0.30       |
| Response-Different | 0.13     | 0.014   | 9.82      | < 0.001*** |
| BlockNum           | -0.047   | 0.0046  | -10.10    | < 0.001*** |
| WordFreq           | -0.0053  | 0.0033  | -1.62     | 0.11       |
| VowelDurationDiff  | 0.26     | 0.15    | 1.69      | 0.091      |
| SpecTiltDiff       | 0.0018   | 0.00070 | 2.63      | 0.0088**   |

Intercept: ContrastType = N; Response = ‘same’

**Table 13** Summary of main results, Exp. 2a

|                              | 'same' responses   | Response Times                 |
|------------------------------|--------------------|--------------------------------|
| Same hph vs. Hph-hph         | not sig.           | not sig.                       |
| Same hph vs. Same non-hph    | higher for non-hph | faster for non-hph             |
| Contrast type N vs. C        | not sig.           | slower for C                   |
| Contrast type N vs. O        | not sig.           | not sig.                       |
| Frequency                    | not sig.           | faster for higher freq         |
| Response spectral tilt diff. | NA                 | slower for 'different'         |
|                              | not sig.           | slower with greater difference |
| Vowel duration diff.         | not sig.           | slower with greater difference |

Experiment 1a and Experiment 2a differed in which acoustic cues were significant predictors of response time, which likely results from differences in production. The mean Euclidian distance between paired items was greater within Experiment 1a, particularly for homophone mates. On the other hand, the other measures exhibited greater distance in Experiment 2a, perhaps due to not being controlled by sentential context.

The main results for Experiment 2a are summarized in Table 13.

### Experiment 2b

#### Methods

**Participants** Twenty-four native speakers of American English (seven male; mean age 21.6) participated in this task and were paid for their participation. Half of the participants also participated in an AX discrimination task (Experiment 1a or Experiment 2a), which they completed prior to this task. The other half of the participants did not participate in any discrimination task.

**Materials** Auditory stimuli were individual English words from Experiment 2a. They were associated with written options that were pairs from Experiment 2a, of the types: (a) homophone-homophone pairs (e.g., *sun-son*); and (b) fillers of phonologically different pairs (e.g., *cat-pat*). The elimination of the third type of contrasts present in Experiment 1b (i.e., *sun-fun*) was aimed at reducing possible fatigue effects generated by the larger number of trials. Given that Experiment 1b found no difference between the condition including this filler type and the condition without them, their exclusion in this experiment is not anticipated to make a difference. Only the results for homophone mates are analyzed, aside from a brief summary of filler results, which confirmed that participants were attending to the task.

**Procedure** Participants heard individual words played in isolation and identified each by selecting one of two written options, following the same procedure as Experiment 1b. As before, the presentation positions of the two items were balanced across participants, to control for possible preference either in use of the arrow keys or in attention to a screen side. There were 240 trials in each block, with 80 trials deciding between homophone mates. The homophone mates are given in the Appendix.

All statistical results are from a mixed effects model, calculated with the lme4 package in R (Bates et al., 2015). *p* values were calculated by the lmerTest package (Kuznetsova et al., 2015). Responses with latencies shorter than 10 ms and the slowest 1% of responses in each category (>4.5 s for homophone mate decisions) were excluded from analysis.

#### Results

Listeners' accuracy for decisions between phonologically distinct words was 96.4%, confirming that they were attending to the task. Responses were slower for decisions about homophone mates (1483 ms) than about phonologically distinct words (1326 ms), though the difference was smaller than it was in Experiment 1b and less consistent across participants.

Table 14 presents a generalized linear mixed effects model including possible factors influencing decisions about homophone mates. The random effects were participant and word pair. The fixed effects were: side of the screen on which the correct answer appeared (left; right); whether or not the participant participated in an AX discrimination task containing homophone stimuli; contrast type of fillers within the block (Onsets; Nuclei; Codas); response time; block number; trial number within the block; log frequency of the acoustically presented item; and log frequency of the competing homophone mate.

**Table 14** glmer model for accuracy in homophone identification

|                    | $\beta$   | SE      | z value | p value   |
|--------------------|-----------|---------|---------|-----------|
| (Intercept)        | 0.25      | 0.21    | 1.21    | 0.23      |
| ScreenSide right   | -0.63     | 0.055   | -11.48  | <0.001*** |
| StimSet no AX task | 0.044     | 0.055   | 0.79    | 0.43      |
| ContrastType C     | -0.045    | 0.067   | -0.68   | 0.50      |
| ContrastType O     | 0.047     | 0.067   | 0.70    | 0.49      |
| ResponseTime       | -0.021    | 0.039   | -0.55   | 0.58      |
| BlockNumb          | 0.039     | 0.034   | 1.15    | 0.25      |
| Trial              | -0.000087 | 0.00040 | -0.22   | 0.83      |
| FreqCorr           | 0.14      | 0.015   | 9.58    | <0.001*** |
| FreqIncorr         | -0.14     | 0.015   | -9.61   | <0.001*** |

Intercept: ScreenSide = left; StimSet = after AX task; ContrastType = N

The by-pair variance and by participant variance were both negligible ( $<0.001$ ).

Accuracy was at chance; the overall mean accuracy was 49.7%. The lack of contrast between hph-hph pairs and pairs of the same homophonic word in Experiment 2a also suggested that the homophone mate pairs in these stimuli have no salient differences. The only strong factors predicting listeners' responses were about response preference, not dependent on the stimuli. This is in contrast to Experiment 1b, in which contrast type and trial number were significant predictors of accuracy, suggesting that they influenced listeners' engagement with the stimuli. The lack of effect of these factors suggests that listeners were not reliably influenced by the form of the stimuli.

As in Experiment 1b, the strongest effect was a preference for selecting the written option on the left side of the screen, so accuracy was higher when the correct answer appeared on the left. Again, stimuli were balanced to have an equal number of correct items appearing on the left and the right side of the screen, so this bias did not influence measurements of perceptually motivated accuracy.

There was no significant effect of the task condition, whether or not this discrimination task followed an AX discrimination task including the same items. There was also no significant effect of trial number or block number on accuracy. All of these null results suggest that prior exposure to the items did not facilitate distinct representations, nor did listeners develop counter-productive strategies of attending to unreliable characteristics based on exposure.

The contrasts present in the phonologically contrastive filler pairs within the block had no effect. This is in contrast with the significant differences between block types in Experiment 1b.

As in Experiment 1b, there was no effect of response time. Response times were faster on average and had a wider range than responses in Experiment 1b, but this does not seem to have been driving any of the results.

As in Experiment 1a, there was a significant effect of frequency on decisions; listeners more frequently selected higher-frequency items as responses, apparent both as a positive effect of the frequency of the stimulus item and a negative effect of the frequency of the homophone mate. The effect in this study was even larger, perhaps because of the larger set of items or because there were fewer other factors influencing decisions.

## Discussion

These experiments are consistent with homophone mates having separate lexical representations and shared phonological representations, though sub-phonological acoustic details can also be weakly associated with particular lexical forms.

Evidence for separate representations needs to be interpreted relative to the representational level that is being accessed by the task. At the lemma level, homophone mates must have distinct representations, given that they have different meanings, though it remains in question whether their phonological representation is shared (cf. Dell, 1990; Jescheniak & Levelt, 1994; Levelt et al., 1999) or separate (cf. Gaskell & Marslen-Wilson, 1997; Seidenberg et al., 1982). Depending on the representation level activated by the task, whether or not a word is lexically ambiguous could produce different response patterns. In tasks that activate competing lexical representations, homophones are identified more slowly than words without homophones (Hino et al., 2002; Pykkänen et al., 2006; Siakaluk et al., 2007), while in tasks that elicit purely phonological or orthographic searches, responses to homophones can be faster than for other words (Borowsky & Masson, 1996; Hino et al., 2002; Kawamoto et al., 1994). In auditory perception tasks, it could be possible to analyze phonological forms without ever activating any word-specific representation, so results may be ambiguous between shared and separate phonological representations.

In Experiment 1a, there was no difference in response times between pairs of the same homophonic word twice (*sun-sun*) and the same non-homophonic word twice (*cat-cat*), which might suggest that listeners were approaching this task phonologically. There was only weak evidence of competition resulting from activation of multiple lemmas, in the marginally larger number of 'same' responses to lexically ambiguous pairs. In comparison, in Experiment 2a, lexically ambiguous phonologically matching forms (*sun-sun*, *sun-son*) were evaluated more slowly and identified as 'different' significantly more often than lexically unambiguous forms (*cat-cat*), suggesting that the lemma level is being activated, producing slower response for items which activate two lemmas than for items which only activate one lemma. The activation of two lemmas creates an additional stage of uncertainty, given that the lexical search cannot be narrowed down to a single item based on the acoustic form. Additional activation at the lemma level also provides a way of explaining the difference between listeners' behavior in the two experiments and is consistent with the slower mean response time in Experiment 2a than in Experiment 1a.

There are several possible explanations for why the two experiments might have elicited different response strategies. A different retrieval pattern for the two tasks might result from the longer duration of words produced in isolation. Longer stimulus durations forced listeners to spend more time on each trial; additional lexical information may have been activated more often in those longer latencies. The difference could also result from difficulty; the stimuli in Experiment 1a, extracted from their original contexts,



were more challenging to identify than the stimuli in Experiment 2b, which were heard in the same context in which they were produced. In more challenging discrimination tasks, ‘different’ responses tend to be slower than ‘same’ responses (Nickerson, 1969); consistent with this effect, responses to the phonologically distinct filler items were slower than responses to phonologically matching pairs in Experiment 1a, but faster than phonologically matching pairs in Experiment 2a. Given the ease of phonological identifications in Experiment 2a, participants might also have interpreted the purpose of the task as lexical, and approached it differently.

There was a small effect of orthography, with most of the difference between hph-hph and same pairs carried by heterographic homophones and not homographic homophones. This may suggest that orthography can have a role in phonological representations, such that it has an influence even in tasks with no orthographic component. Most studies on homophones have not found an effect of spelling (e.g., Biedermann & Nickels, 2008; Caramazza et al., 2001). However, there is often a contrast between semantically unrelated homophones and the related meanings of polysemous words (e.g., Klepousniotou & Baum, 2007; Pyllkkänen et al., 2006; Rodd et al., 2002), so the small effect in this experiment might suggest that some of the homographic homophones would actually be better categorized as polysemously related forms.

The relationship between response time and lexical frequency can inform the representation of homophone mates. There is a substantial amount of work demonstrating processing advantages for high frequency words, mostly in production and comprehension of orthographic stimuli, including lexical decision (Murray & Forster, 2004; Stanners, Jastrzembki, & Westbrook, 1975), semantic categorization (Lewellen, Goldinger, Pisoni, & Greene, 1993; Monsell, Doyle, & Haggard, 1989), picture naming (Carroll & White, 1973; Oldfield & Wingfield, 1965), and fixation duration in reading (Rayner & Duffy, 1986). Most studies with homophones have found that the measure of homophone frequency that best fits with these patterns is the individual frequency of each homophone mate (e.g., Caramazza et al., 2001; Simpson & Burgess, 1985).

Separate lexical frequency was apparent from frequency effects in identification decisions in Experiments 1b and 2b; the higher frequency item was more frequently selected as an answer. This effect can be attributed to stronger activation for the orthographic form associated with the higher frequency semantic entry. It does not necessarily indicate what activation would be elicited by the ambiguous acoustic form in the absence of orthographic forms. The slower response times and greater numbers of ‘different’ responses to lexically ambiguous pairs in the discrimination tasks in Experiment 2a would be expected if both forms are being activated in these tasks too; all of these

experiments provide evidence for a lexical task strategy and do not indicate whether phonological representations of homophone mates are shared or not.

In Experiment 2a, word frequency was a significant predictor of accuracy and response time, but the effect was driven by lexically unambiguous items, and was eliminated when the models excluded those items. These results are consistent with previous work demonstrating that higher frequency words are activated more quickly than lower frequency words (e.g., Binder & Rayner, 1998; Simpson & Burgess, 1985). The results of this experiment are consistent with stronger activation of higher frequency words, which allows listeners to more quickly and more accurately recognize their phonological forms and map two such forms onto the same lexical representation. The lack of frequency effect within homophones could suggest that lexical ambiguity interferes with this mapping.

The lack of frequency effect in Experiment 1a might result from limitations in the frequency distribution of the stimuli. However, despite the large amount of work demonstrating frequency effects in production and in perception of orthographic stimuli, there is less work demonstrating this effect in acoustic perception; some studies have found that listeners are more likely to interpret ambiguous stimuli as matching higher frequency words, e.g., for identification of words in noise (Howes, 1957) or with phonologically ambiguous segments (Connine et al., 1993), though Samuel (1981) did not find an effect of word frequency in phoneme restoration. Lexical frequency effects occur slightly later in processing than phonological effects, as is seen in MEG responses to orthographic words, both non-homophones (Pyllkkänen et al., 2002) and homophones (Simon et al., 2012). Given the evidence for largely phonological search strategies in this experiment, such that activation of the lexical level is never fully reached, the results might suggest that lexical frequency is not a good test for representations at the phonological level.

Homophone mates can exhibit significant acoustic differences in production, based on factors such as lexical frequency (Gahl, 2008; Guion, 1995) and part of speech (Conwell, 2017; Sorensen et al., 1978). If homophone mates are acoustically identical or if listeners are not sensitive to what differences might exist, the discrimination tasks should exhibit the same patterns of responses to pairs of homophone mates (e.g., *sun-son*) and pairs with the same homophone twice (e.g., *sun-sun*). The different contexts of production, in isolation or extracted from sentences, additionally test how production environment influences the acoustic form of stimuli, including differences that listeners may be sensitive to.

In Study 1, with stimuli produced in sentences, there was greater acoustic distance between different speakers’ productions of homophone mates than between different speakers’ productions of the same word, in Euclidean

distance between vowels, vowel duration, F0 maximum, and spectral tilt, though the differences did not reach significance. In contrast, in Study 2, with stimuli produced in isolation, there was no such trend. These results are consistent with Guion's (1995) finding that differences in production are only apparent within contexts that activate the meaning of the word. The results from Study 1 should be interpreted cautiously, as hph-hph pairs had slight differences in their production environments, while tokens for same pairs came from identical sentences, which might explain the greater differences in the former category, though sentences were selected to be similar in duration and prosodic structure, to minimize such effects.

However, the above chance accuracy in Experiment 1b suggests that the acoustic differences were relevant to the lexical items. That is, even if the differences are effects of the sentential environments, those characteristics are associated with the lexical items; this association could result from different patterns of the environments which each word most frequently appears in. While accuracy in homophone identification could be interpreted as supporting separate phonological representations of homophone mates, which include acoustic details based on patterns of usage (cf. Gahl, 2008; Guion, 1995), the weakness of listeners' perceptual sensitivity to these differences opposes this interpretation. The results of this study nevertheless suggest some association of phonetic details with particular lexical items (cf. Johnson, 1997; Pierrehumbert, 2002).

If differences between homophone mates can be used categorically, these pairs should receive a larger number of 'different' responses than other pairs. However, pairs of homophone mates (*sun-son*) did not exhibit significantly more 'different' responses than same pairs of words with homophones (*sun-sun*). This lack of difference, in comparison to listeners' high accuracy for identifying phonologically distinct items as 'different', indicates that these acoustic differences between homophone mates are not treated as contrastive, which is consistent with the absence of these differences when homophones are produced in isolation.

The lack of acoustic differences in Study 2 suggests that homophone mates do not have separate phonological representations. If the acoustic differences in Study 1 were an inherent part of the phonological representations of the homophone mates, they should be present regardless of the environment of production. The absence of these differences when words are produced in isolation suggests that they result from pressures of context in production such as phrasal position (Conwell, 2017; Sorensen et al., 1978) and predictability of the word based on context (Jurafsky et al., 2002; Scarborough, 2010). Listeners can be sensitive to subphonemic details (cf. Andruski et al., 1994; Pisoni & Tash, 1974), including cues for grammatical category (Conwell & Morgan, 2012), so effects of the acoustic

differences on responses in Study 1 do not prove that these differences are phonological.

The different response times across pair types in the discrimination tasks indicate a sensitivity to acoustic differences. Listeners' response times for the different pair types differed between the two sets of stimuli, consistent with the different acoustic patterns of the stimuli. In Experiment 1a, with stimuli produced in sentential contexts, pairs of homophone mates were evaluated more slowly than same pairs of words with homophones, suggesting that listeners are sensitive to the differences between items. In Experiment 2a, with stimuli produced in isolation, homophone mates did not exhibit longer latencies than pairs of the same homophonic word.

In both experiments, at least some of the acoustic measures were significant predictors of response time and others exhibited similar trends, though the experiments differed in the strength of each measure. Given this evidence for listeners being influenced by acoustic distance even within phonological categories, it follows that a pair type with overall greater distance would also exhibit longer response times.

Identification decisions with pairs of homophone mates in Experiments 1b and 2b provided another measure of listeners' sensitivity to acoustic differences. Listeners were significantly above chance at identifying homophones in Experiment 1b, with stimuli which were extracted from meaningful sentences, though this accuracy was very low relative to their ability to discriminate between regular phonological contrasts. In Experiment 2b, the stimuli produced in isolation did not exhibit this pattern. These results suggest that listeners are weakly capable of discriminating between homophone mates, as long as they include the acoustic differences that are induced by their syntactic and semantic context, which are not produced when items are read in isolation.

Previous work has not found above chance accuracy for distinguishing between homophone mates, though listeners' response preferences for some particular items suggests that they can be sensitive to acoustic details (Bond, 1973); that lack of discrimination might reflect effects of how stimuli were elicited, most crucially depending on whether the stimuli came from meaningful sentences or from productions in isolation. The additional significant factors within Experiment 1b also indicate some of the ways that methods of stimulus presentation influence homophone discrimination. Accuracy was lower in blocks in which filler trials contained items differing by coda than in blocks in which fillers differed in onsets or nuclei, which might suggest that directing listeners' attention to different parts of the syllable can influence whether they focus on acoustic cues which align with actual differences between homophone mates, and indicates the importance of filler

selection. Total absence of filler trials might have a still different effect; this study did not test such a condition, but it is likely to be worth investigating in future work.

In the discrimination tasks, the three different block types based on the filler trials (differences in onsets, nuclei, or codas) tested whether these filler trials set up expectations about which parts of the words could contain differences. Previous work has established that listening context is important in setting up listeners' expectations and influencing their strategies in completing tasks; changing filler items can influence response patterns to the target items (e.g., Borowsky & Masson, 1996; Grainger et al., 2001; Vitevitch & Luce, 1999).

Differences between the block types demonstrate the effect of expectations on processing; even in phonologically matching pairs, response times reflected where contrasts in the phonologically different pairs in that block were, suggesting that listeners were sensitive to how much of the word they needed to hear in order to determine whether or not there was a contrast. However, there was only a significant difference between coda blocks and other blocks, not between onset and nucleus blocks. These effects are particularly relevant for informing work using acoustic discrimination tasks, in which the form of the fillers is not always strictly controlled.

Given that there were differences in response time based on the type of fillers in the block, the relationship with pair type could establish if there are certain contrasts, based on the sounds or their positions, which listeners were more uncertain about than others. In particular, an interaction could establish if listeners were attending to particular subparts of homophone mates as the positions of potential contrasts that might allow them to discriminate between those homophone mates. However, there was no interaction between pair type and contrast type, so there is no evidence that listeners expect certain parts of a word to contain contrasts between homophone mates.

## Conclusions

This set of experiments provides data for lexical access of homophones based on acoustic input, whereas much of the previous literature on homophone storage is based on orthographic input. Results suggest that both lexical entries are activated by acoustically presented homophones in some listening contexts, though listening tasks that are approached phonologically might not be activating lexical forms at all.

Although listeners could decide between homophone mates with slightly better than chance accuracy, the consistency of 'same' responses to these pairs in the discrimination

tasks indicates that the differences are non-contrastive. The absence of consistent phonetic differences between homophone mates across different elicitation environments further suggests that the details are not inherent to each form. The results are best explained by homophone mates having shared phonological representations, though listeners may weakly associate phonetic details with particular items based on their typical realizations in context.

The environment of elicitation influenced the stimuli and subsequently listeners' responses to them; greater acoustic distance between homophone mates than between lexically matching items was only apparent for words produced in sentences, and not for words produced in isolation. Acoustic distance in several measures was a predictor of response time in the discrimination tasks. This indicates that listeners are sensitive to subphonemic details, even when their decision patterns are categorical.

**Acknowledgements** I would like to thank Shu-hao Shih for asking the question that inspired this project.

## Appendix

**Table 15** Homophone mates, Study 1

| Homographic                                | Heterographic         |
|--|-----------------------|
| back (backwards) - back (rear)             | bear (animal) - bare  |
| bank (institution) - bank (river)          | break - brake         |
| bar (tavern) - bar (prohibit)              | creek - creak         |
| bark (dog) - bark (tree)                   | dough - doe           |
| bat (sports) - bat (animal)                | fair - fare           |
| can (container) - can (be able)            | halve - have          |
| die (cube) - die (pass away)               | heard - herd          |
| dress (gown) - dress (clothe)              | made - maid           |
| file (organize) - file (tool)              | mail - male           |
| fine (fee) - fine (good)                   | meat - meet           |
| ground (earth) - ground (powdered)         | night - knight        |
| light (lightweight) - light (illumination) | pale - pail           |
| like (fond) - like (similar)               | pear - pair           |
| match (correspond) - match (fire)          | peer - pier           |
| mind (brain) - mind (care)                 | piece - peace         |
| nail (finger) - nail (construction)        | read (past) - red     |
| park (stop) - park (garden)                | read (present) - reed |
| plant (verb) - plant (noun)                | sole - soul           |
| rock (stone) - rock (sway)                 | steak - stake         |
| sign (marker) - sign (endorse)             | steal - steel         |
| sink (descend) - sink (wash basin)         | sun - son             |
| type (write) - type (kind)                 | waist - waste         |
| wave (ripple) - wave (gesture)             | write - right         |

**Table 16** Homophone mates, Study 2

|                 |                |                |               |
|-----------------|----------------|----------------|---------------|
| ad - add        | groan - grown  | plain - plane  | sole - soul   |
| ail - ale       | heard - herd   | plait - plate  | son - sun     |
| brake - break   | hole - whole   | raised - razed | suite - sweet |
| chord - cord    | knight - night | right - write  | vain - vein   |
| coarse - course | lacks - lax    | sail - sale    | vial - vile   |
| crews - cruise  | loot - lute    | scene - seen   | wail - whale  |
| doe - dough     | made - maid    | sea - see      | wait - weight |
| feat - feet     | mail - male    | shear - sheer  | weak - week   |
| flea - flee     | meat - meet    | sight - site   | wet - whet    |
| foul - fowl     | peace - piece  | soar - sore    | whine - wine  |

## References

- Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52(3), 163–187.
- Antón-Méndez, I., Schütze, C. T., Champion, M. K., & Gollan, T. H. (2012). What the tip-of-the-tongue (tot) says about homophone frequency inheritance. *Memory & Cognition*, 40(5), 802–811.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research*, 24(1), 57–65.
- Biedermann, B., & Nickels, L. (2008). Homographic and heterographic homophones in speech production: Does orthography matter? *Cortex*, 44(6), 683–697.
- Binder, K. S., & Rayner, K. (1998). Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading. *Psychonomic Bulletin & Review*, 5(2), 271–276.
- Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [computer program]. version 6.0.30. <http://www.praat.org/>
- Bond, Z. S. (1973). The perception of sub-phonemic phonetic differences. *Language and Speech*, 16(4), 351–355.
- Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 63.
- Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(6), 1430–1450.
- Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *The Quarterly Journal of Experimental Psychology*, 25(1), 85–95.
- Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(1), 81–94.
- Conwell, E. (2017). Prosodic disambiguation of noun/verb homophones in child-directed speech. *Journal of Child Language*, 44(3), 734–751.
- Conwell, E., & Morgan, J. L. (2012). Is it a noun or is it a verb? Resolving the ambigatogicality problem. *Language Learning and Development*, 8(2), 87–112.
- Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J. T. (1978). Phonological recoding and lexical access. *Memory & Cognition*, 6(4), 391–402.
- Davies, M. (2008). The corpus of contemporary American English (COCA): 560 million words, 1990–present. <https://corpus.byu.edu/coca/>.
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5(4), 313–349.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27(4), 429–446.
- Folk, J. R., & Morris, R. K. (1995). Multiple lexical codes in reading: Evidence from eye movements, naming time, and oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1412–1429.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3), 474–496.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12(5–6), 613–656.
- Grainger, J., Van Kang, M. N., & Segui, J. (2001). Cross-modal repetition priming of heterographic homophones. *Memory & Cognition*, 29(1), 53–61.
- Guion, S. G. (1995). Word frequency effects among homonyms. In *Texas Linguistic Forum*, (Vol. 35, pp. 103–116).
- Hall, K. C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review*, 30(2), 215–275.
- Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 686–713.
- Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America*, 29(2), 296–305.
- Jastrzemski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology*, 13(2), 278–305.
- Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 824–843.
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego: Academic Press.
- Jurafsky, D., Bell, A., & Girand, C. (2002). The role of the lemma in form variation. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 3–34). Berlin: Mouton de Gruyter.
- Kawamoto, A. H., Farrar, W. T., & Kello, C. T. (1994). When two meanings are better than one: Modeling the ambiguity advantage using a recurrent distributed network. *Journal of Experimental Psychology: Human Perception and Performance*, 20(6), 1233–1247.
- Kellas, G., Ferraro, F. R., & Simpson, G. B. (1988). Lexical ambiguity and the timecourse of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 14(4), 601–609.
- Klepousiotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, 20(1), 1–24.
- Klepousiotou, E., Pike, G. B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy. *Brain and Language*, 123(1), 11–21.

- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2015). ImerTest: Tests in linear mixed effects models. R package version 2.0-29.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38.
- Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General*, 122(3), 316–330.
- Lohman, A. (2017). Cut(n) and cut(v) are not homophones: Lemma frequency affects the duration of noun-verb conversion pairs. *Journal of Linguistics*, 1–25.
- Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is initially phonological: I. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General*, 123(2), 107–128.
- Masson, M. E., & Freedman, L. (1990). Fluent identification of repeated words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3), 355–373.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), B33–B42.
- McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 621–638.
- Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, 118(1), 43.
- Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, 111(3), 721–756.
- Nickerson, R. S. (1969). ‘same’-‘different’ response times: A model and a preliminary test. *Acta Psychologica*, 30(Attention and Performance II), 257–275.
- Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4), 273–281.
- Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9(3), 225–236.
- Pierce, J. W. (2007). PsychoPy–Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13.
- Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 101–140). Berlin: Mouton de Gruyter.
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2), 285–290.
- Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience*, 18(1), 97–109.
- Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language*, 81(1–3), 666–678.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3), 191–201.
- Renwick, M. E. L., & Ladd, D. R. (2016). Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrasts. *Laboratory Phonology*, 7(1), 19.
- Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2), 245–266.
- Samuel, A. G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 7(5), 1124.
- Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. *Laboratory Phonology*, 10, 557–586.
- Schvaneveldt, R. W., Meyer, D. E., & Becker, C. A. (1976). Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 2(2), 243–256.
- Scobbie, J. M., & Stuart-Smith, J. (2008). Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. In P. Avery, B. E. Dresher, & K. Rice (Eds.), *Contrast in Phonology: Theory, Perception, Acquisition* (pp. 87–113). Berlin: de Gruyter.
- Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, 14(4), 489–537.
- Siakaluk, P. D., Pexman, P. M., Sears, C. R., & Owen, W. J. (2007). Multiple meanings are not necessarily a disadvantage in semantic processing: Evidence from homophone effects in semantic categorisation. *Language and Cognitive Processes*, 22(3), 453–467.
- Simon, D. A., Lewis, G., & Marantz, A. (2012). Disambiguating form and lexical frequency effects in MEG responses using homonyms. *Language and Cognitive Processes*, 27(2), 275–287.
- Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, 11(1), 28–39.
- Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, 30(6), 627–643.
- Sorensen, J. M., Cooper, W. E., & Paccia, J. M. (1978). Speech timing of grammatical categories. *Cognition*, 6(2), 135–153.
- Stanners, R. F., Jastrzemski, J. E., & Westbrook, A. (1975). Frequency and visual quality in a word-nonword classification task. *Journal of Verbal Learning and Verbal Behavior*, 14(3), 259–264.
- Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, 18(4), 427–440.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40(3), 374–408.
- Wheeldon, L. R., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *The Quarterly Journal of Experimental Psychology*, 44(4), 723–761.