frontiers in
PSYCHOLOGY

METHODS ARTICLE
published: 15 March 2012
doi: 10.3389/fpsyg.2012.00049

A collection of pseudo-words to study multi-talker speech intelligibility without shifts of spatial attention

***Kachina Allen*[1], *David Alais*[2]* *and Simon Carlile*[3]**

[1] *Auditory, Brain and Cognitive Development Laboratory, McGill University, Montreal, QC, Canada*
[2] *School of Psychology, University of Sydney, Sydney, NSW, Australia*
[3] *Discipline of Physiology, School of Medical Sciences, University of Sydney, Sydney, NSW, Australia*

Edited by:
Claude Alain, Rotman Research Institute, Canada

Reviewed by:
Gregg H. Recanzone, University of California, USA
Kimmo Alho, University of Helsinki, Finland

***Correspondence:**
David Alais, School of Psychology, University of Sydney, 506 Griffith Taylor Building, Sydney, NSW 2006, Australia.
e-mail: david.alais@sydney.edu.au

A new collection of pseudo-words was recorded from a single female speaker of American English for use in multi-talker speech intelligibility research. The pseudo-words (known as the KARG collection) consist of three groups of single syllable pseudo-words varying only by the initial phoneme. The KARG method allows speech intelligibility to be studied free of the influence of shifts of spatial attention from one loudspeaker location to another in multi-talker contexts. To achieve this, all KARG pseudo-words share the same concluding rimes, with only the first phoneme serving as a distinguishing identifier. This ensures that listeners are unable to correctly identify the target pseudo-word without hearing the initial phoneme. As the duration of all the initial phonemes are brief, much shorter than the time required to spatially shift attention, the KARG method assesses speech intelligibility without the confound of shifting spatial attention. The KARG collection is available free for research purposes.

Keywords: speech, spatial attention, multiple talkers

INTRODUCTION

Previously available corpora used in multi-talker speech intelligibility trials have included the coordinate response measure (CRM; Bolia et al., 2000), the British Bamford–Kowal–Bench sentences (Bench and Bamford, 1979), the SPIN lists (Kalikow et al., 1977), and the Modified Rhyme Test (House et al., 1963). These involve a carrier phrase or sentence and identifying words within them that are used for scoring speech intelligibility. These corpora are useful because they simulate the real-world listening environment where keywords are usually embedded within a lengthy speech stream. Their long duration, however, means these corpora are not well suited to speech intelligibility research when multiple talkers at different spatial locations are involved because rapid shifts in spatial attention may confound the results. Here we present a short duration method which overcomes this problem and allows spatial aspects of multi-talker speech intelligibility to be studied without the influence of spatial attention.

It has been estimated that the initial focusing of auditory attention on a spatial location can be done in as little as 80 ms (Teder-Sälejärvi and Hillyard, 1998), and other studies show that shifting attention laterally from one location to another may require 200 ms or so (Treisman, 1971; Massaro, 1976). Thus a discourse or carrier phrase longer than this temporal window would allow the listener time to move their focus of attention to the speech location prior to the scoring words. Even single words could provide a listener with enough time to shift their attention to the location of the speech signal and infer the entire word from the final phoneme, especially if the word-set were limited. The pseudo-words described here (the KARG collection) use only the first phoneme of non-words as identifiers, with identical concluding rimes to ensure that the subject is unable to identify the target pseudo-word without hearing the initial word-sound.

Another reason for using pseudo-words is to avoid any type of interference that word familiarity may have on recognition. The frequency with which a word is used in regular speech usage has been demonstrated to affect response times and accuracy, with higher rates of word usage leading to better performance (Luce and Pisoni, 1998; Connine, 2004). The KARG collection consists entirely of pseudo-words and therefore is not subject to word frequency confounds.

Unlike other corpora where the target is identified by a preceding key word or by a talker, the target in the KARG collection is identified by the concluding sound alone, after the initial phoneme. This prevents the listener from switching their attention to the target location before the scoring portion of the word is presented. To maximize reliance on spatial cues, the same talker was used to record the set of masker and target pseudo-words thus ensuring that cues such as gender, pitch, or *f*0 could not be used to identify the target. Use of the same talker for target and masker not only maximizes spatial release from masking (e.g., Festen and Plomp, 1990; Noble and Perrett, 2002), it also maximizes the amount of informational masking in the pseudo-words (Brungart, 2001).

The KARG recordings cover a broad frequency spectrum. This is important as the ability to locate sound sources is significantly reduced when bandpassed stimuli are used (Middlebrooks, 1992). Carlile et al. (1999) demonstrated a significant reduction in localization accuracy by human subjects for 2 kHz low-pass noise stimuli. Similar reductions also occur for 8 kHz low-passed speech stimuli (Best et al., 2005). Langendijk and Bronkhorst (2002) argued that the 6- to 12-kHz spectrum provides cues for

up-down localization, while the 8- to 16-kHz band is important for front-back resolution. Many of the available corpora, including the commonly used CRM corpus mentioned above, are low-pass filtered at 8 kHz (Bolia et al., 2000) which unfortunately deprives the listener of useful high-frequency cues to the talker's spatial location that would normally be available in the transient elements of speech, particularly fricatives, and plosives. To preserve these high-frequency cues to talker location, the pseudo-words described here were recorded using a broad bandwidth (0–22.5 kHz).

The KARG collection has been designed to allow research into multi-talker speech intelligibility without the confounding effects of auditory spatial attention. Many previous studies of auditory attention have used noise-burst or tone stimuli (Teder and Näätänen, 1994; Teder-Sälejärvi and Hillyard, 1998; Teder-Sälejärvi et al., 1999; Widmann and Schröger, 1999; Sach et al., 2000; Sach and Bailey, 2004). The KARG collection will allow easy comparison of focused versus non-focused attention conditions and provide a useful paradigm for understanding speech phenomena such as the cocktail-party problem (Cherry, 1953; Bronkhorst, 2000) without the confound of unwanted spatial attentional shifts (Allen et al., 2008, 2009).

THE "KARG" COLLECTION

The full KARG collection consists of five recordings of each of 18 pseudo-words. The following three word endings – "-arg," "-org," and "-oog" – were each combination six different consonants – "p," "t," "k," "b," "d," "g" – to create 18 pseudo-words. The pseudo-words are designed to have a target, consisting of one of the voiceless consonants "p," "t," or "k," followed by a concluding sound for identification ("-arg," "-org," or "-oog"). To maximize masking of the initial phoneme, interferers beginning with the voiced consonants "b," "d," or "g" are used, followed by an alternative word ending ("-arg," "-org," or "-oog").

Five recordings of each pseudo-word are contained in the KARG collection. This is designed to overcome the problem that with repeated trials, small differences in recording, or articulation of the last part of a given pseudo-word may allow a listener to recognize that pseudo-word and thus correctly identify the initial phoneme. By randomly selecting between the five recordings of each pseudo-word, this problem is overcome.

To ensure that the KARG collection is useful for controlling auditory spatial attention shifts, the scoring section of the pseudo-word must occur prior to the listener being able to refocus attention at a new location. The length of the identification segment (the initial phoneme) was thus calculated for all pseudo-words in the KARG collection. All of the identifying initial consonants were less than 80 ms in duration and the average over all recordings of all pseudo-words was 60.2 ± 14.1 ms. This is less than the 80-ms minimum that Teder-Sälejärvi and Hillyard (1998) suggested was required for re-deploying spatial attention to a new location.

Figure 1 gives a summary of the durations of the initial phonemes. Each of the six initial consonants is listed separately with the three possible endings "-arg" (black columns), "-oog" (white columns), and "-org" (gray columns). The data columns indicate the average length of the phonemes for the five recordings of each pseudo-word. Error bars showing SEM are included. The dashed horizontal line shows the 80 ms minimum for re-deploying spatial attention to a new location (Teder-Sälejärvi and Hillyard, 1998). For all KARG pseudo-words, the initial consonants are all below this time limit.

METHOD OF COLLECTION

A single female speaker of American English was used to record all pseudo-words in the KARG collection. All recording was done in a sound-attenuated audiometric booth (3.5 m × 4.6 m × 2.4 m), lined with 3″ acoustic foam to create a semi-anechoic environment

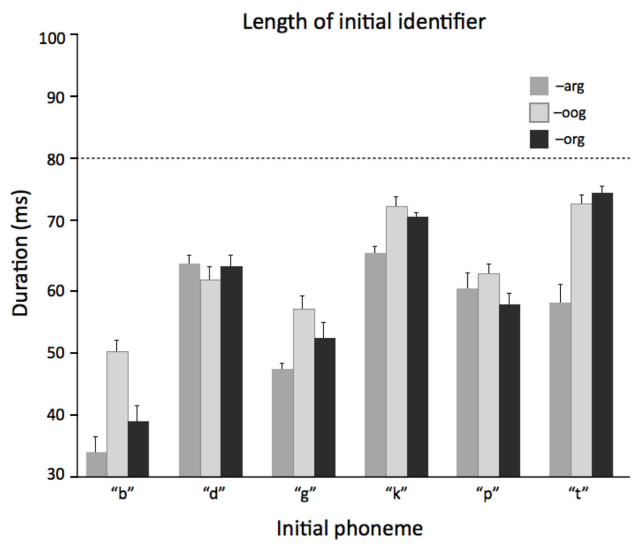

FIGURE 1 | The mean duration of the initial identifying sound for the group of five recordings of each of the possible pseudo-words. The dashed horizontal line at 80 ms denotes the minimum time required to re-orient auditory spatial attention, as estimated by Teder-Sälejärvi and Hillyard (1998). All initial identifying sounds are less than 80 ms in duration. Error bars are SE.

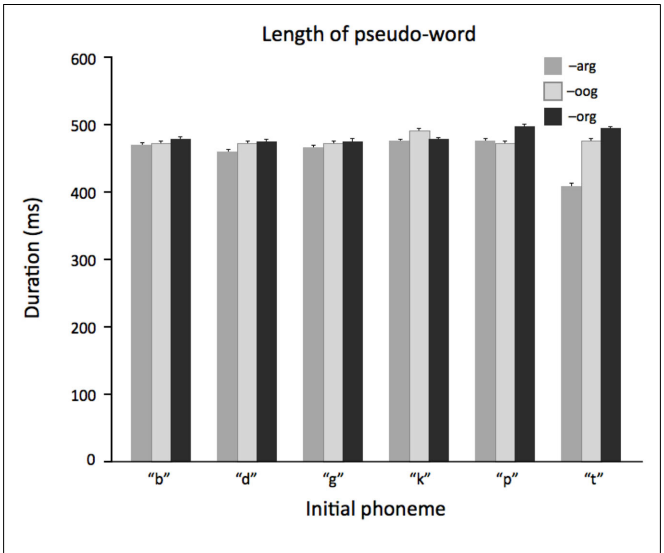

FIGURE 2 | The mean duration in milliseconds of the five recordings of each pseudo-word. Error bars are SE.

suitable for speech recording. Recordings were carried out in a single session, with the talker seated 5 cm from a Shure SM48 microphone. The sound was collected using the CSL KAY 4500 audiometric collection software at a sampling rate of 44100 kHz.

After practice to ensure consistency of pronunciation, the talker was presented with a list of written pseudo-words. She was instructed to speak each word on the list a minimum of 10 times in a clear and natural voice. The recording was collected as a single file. To permit synchronized playback, each pseudo-word was edited from the recording with all intervening silence removed (using the freely available Audacity music software v1.2.4: http://audacity.sourceforge.net/).

For each pseudo-word, all 10 recordings were aurally screened for audio quality and the five that were the clearest and most similar in duration to others were selected for inclusion in the KARG collection. Matlab software (The Mathworks, Inc.) was used to scale each recording so that all had the same root-mean squared average power. The overall duration of each of the 18 pseudo-words in the KARG collection is shown in **Figure 2**, which shows the mean of the five recordings of each pseudo-work with error bars showing 1 SEM. The grand mean duration of all KARG pseudo-words is 476.1 ± 24 ms.

AVAILABILITY OF THE KARG COLLECTION

The KARG collection is available free of charge to researchers. Contact the author for further details.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council. Thanks to Karen Froud and Teacher's College Columbia University for the use of facilities to record and manipulate the stimuli. Thanks also to the anonymous student who recorded the corpus.

REFERENCES

Allen, K., Alais, D., and Carlile, S. (2008). The contribution of talker characteristics and spatial location to auditory streaming. *J. Acoust. Soc. Am.* 123, 1562–1570.

Allen, K., Alais, D., and Carlile, S. (2009). Speech intelligibility reduces over distance from an attended location: evidence for an auditory spatial gradient of attention. *Atten. Percept. Psychophys.* 71, 164–173.

Bench, J., and Bamford, J. M. (eds). (1979). *Speech Hearing Tests and the Spoken Language of Hearing-Impaired Children.* London: Academic Press.

Best, V., Carlile, S., Jin, C., and van Schaik, A. (2005). The role of high frequencies in speech localization. *J. Acoust. Soc. Am.* 118, 353–363.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). A speech corpus for multitalker communications research. *J. Acoust. Soc. Am.* 107, 1065–1066.

Bronkhorst, A. (2000). The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acta Acust.* 86, 117–128.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* 109, 1101–1109.

Carlile, S., Delaney, S., and Corderoy, A. (1999). The localisation of spectrally restricted sounds by human listeners. *Hear. Res.* 128, 175–189.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979.

Connine, C. M. (2004). It's not what you hear but how often you hear it: on the neglected role of phonological variant frequency in auditory word recognition. *Psychon. Bull. Rev.* 11, 1084–1089.

Festen, J. M., and Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J. Acoust. Soc. Am.* 88, 1725–1736.

House, A. S., Williams, C., Hecker, M. H., and Kryter, K. D. (1963). Psychoacoustic speech tests: a modified rhyme test. Techn Docum Rep Esd-Tdr-63-403. *Tech. Doc. Rep. U. S. Air. Force Syst. Command. Electron. Syst. Div.* 86, 1–44.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J. Acoust. Soc. Am.* 61, 1337–1351.

Langendijk, E. H., and Bronkhorst, A. W. (2002). Contribution of spectral cues to human sound localization. *J. Acoust. Soc. Am.* 112, 1583–1596.

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19, 1–36.

Massaro, D. W. (1976). Perceiving and counting sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 337–346.

Middlebrooks, J. C. (1992). Narrow-band sound localization related to external ear acoustics. *J. Acoust. Soc. Am.* 92, 2607–2624.

Noble, W., and Perrett, S. (2002). Hearing speech against spatially separate competing speech versus competing noise. *Percept. Psychophys.* 64, 1325–1336.

Sach, A., Hill, N., and Bailey, P. (2000). Auditory spatial attention using interaural time differences. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 717–729.

Sach, A. J., and Bailey, P. J. (2004). Some characteristics of auditory spatial attention revealed using rhythmic masking release. *Percept. Psychophys* 66, 1379–1387.

Teder, W., and Näätänen, R. (1994). Event-related potentials demonstrate a narrow focus of auditory spatial attention. *Neuroreport* 5, 709–711.

Teder-Sälejärvi, W., and Hillyard, S. (1998). The gradient of spatial auditory attention in free field: an event-related potential study. *Percept. Psychophys.* 60, 1228–1242.

Teder-Sälejärvi, W., Hillyard, S., Roder, B., and Neville, H. (1999). Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials. *Brain Res. Cogn. Brain Res.* 8, 213–227.

Treisman, A. M. (1971). Shifting attention between the ears. *Q. J. Exp. Psychol.* 23, 157–167.

Widmann, A., and Schröger, E. (1999). "ERP indications for sustained and transient auditory attention with different lateralization cues," in *Psychophysics, Physiology, and Models of Hearing*, eds V. Hohmann, B. Kollmeier, and T. Dau (Singapore: World Scientific), 47–50.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 October 2011; accepted: 08 February 2012; published online: 15 March 2012.

*Citation: Allen K, Alais D and Carlile S (2012) A collection of pseudo-words to study multi-talker speech intelligibility without shifts of spatial attention. Front. Psychology **3**:49. doi: 10.3389/fpsyg.2012.00049*

This article was submitted to Frontiers in Auditory Cognitive Neuroscience, a specialty of Frontiers in Psychology.