Aero-tactile integration during speech perception: Effect of response and stimulus characteristics on syllable identification

Donald Derrick,[1,a)] Jilcy Madappallimattam,[2] and Catherine Theys[2]
[1]New Zealand Institute of Language, Brain, and Behaviour, University of Canterbury, 20 Kirkwood Avenue, Upper Riccarton, Christchurch 8041, New Zealand
[2]School of Psychology, Speech and Hearing, University of Canterbury, 20 Kirkwood Avenue, Upper Riccarton, Christchurch 8041, New Zealand

(Received 14 February 2019; revised 14 August 2019; accepted 16 August 2019; published online 11 September 2019)

Integration of auditory and aero-tactile information during speech perception has been documented during two-way closed-choice syllable classification tasks [Gick and Derrick (2009). Nature **462**, 502–504], but not during an open-choice task using continuous speech perception [Derrick, O'Beirne, Gorden, De Rybel, Fiasson, and Hay (2016). J. Acoust. Soc. Am. **140**(4), 3225]. This study was designed to compare audio-tactile integration during open-choice perception of individual syllables. In addition, this study aimed to compare the effects of place and manner of articulation. Thirty-four untrained participants identified syllables in both auditory-only and audio-tactile conditions in an open-choice paradigm. In addition, forty participants performed a closed-choice perception experiment to allow direct comparison between these two response-type paradigms. Adaptive staircases, as noted by Watson [(1983). Percept. Psychophys. **33**(2), 113–120] were used to identify the signal-to-noise ratio for identification accuracy thresholds. The results showed no significant effect of air flow on syllable identification accuracy during the open-choice task, but found a bias towards voiceless identification of labials, and towards voiced identification of velars. Comparison of the open-choice results to those of the closed-choice task show a significant difference between both response types, with audio-tactile integration shown in the closed-choice task, but not in the open-choice task. These results suggest that aero-tactile enhancement of speech perception is dependent on response type demands. © *2019 Acoustical Society of America.*

https://doi.org/10.1121/1.5125131

[BVT] Pages: 1605–1614

I. INTRODUCTION

To understand speech, we do not just use auditory information, but also information from other senses. Integration of the auditory signal with air flow directed at the skin (hereafter *tactile*) has been well documented (Derrick and Gick, 2013; Fowler and Dekle, 1991; Gick and Derrick, 2009; Goldenberg *et al.*, 2015; Treille *et al.*, 2014). For instance, congruent presentation of air flow can enhance accuracy of two-way forced-choice (2AFC) identification of voiceless stop onset syllables from about 68.6% to 76.9% (8.3% range) when applied to the suprasternal notch (neck) (Gick and Derrick, 2009). Such results are similar to enhancement shown in audio-visual integration decades before (McGurk and MacDonald, 1976; Sumby and Pollack, 1954), at least in simple 2AFC paradigms.

Audio and tactile speech stimuli do not need to be well-related in space—an air puff to the ankle will enhance speech perception almost as well as air puffs to the head, neck, or hand (Derrick and Gick, 2013). However, tactile stimuli must otherwise be appropriate to the speech act. Air flow contacting skin integrates with auditory information to affect speech perception but taps on the skin do not (Gick and Derrick, 2009). It has also been shown that audio and tactile stimuli need to be temporally aligned. Perceivers benefit most when the air flow occurs during or shortly after the relevant speech auditory (Gick *et al.*, 2010) or visual (Bicevskis *et al.*, 2016) signal.

In addition, absence or presence of airflow aids differentiation between aspirated and unaspirated syllables (e.g., /pa/ vs /ba/; /ta/ vs /da/), but airflow does not enhance differentiation between two similarly fricated speech sounds (e.g., /d͡ʒa/ vs /da/; /t͡ʃa/ vs /ta/; /t͡ʃa/ vs /ʃa/) (Derrick *et al.*, 2014b). This outcome suggests that the influence of tactile information on speech perception may be specific to the stimulus characteristics of the syllables used.

Audio-tactile integration has been shown in studies focusing on syllable identification (Derrick and Gick, 2013; Gick and Derrick, 2009). This has been extended to monosyllabic word identification (Derrick *et al.*, 2019), but not to integration in open-choice experiments involving continuous speech perception of multiple words (Derrick *et al.*, 2016). There are at least two possible underlying factors that may explain the difference between the syllable-based findings and the more recent continuous speech perception results. These include differences in (1) response type (open- vs closed-choice), and (2) stimulus type (syllables vs sentences).

A. Response type differences

One major difference between the original study showing audio-tactile integration (Gick and Derrick, 2009) and the continuous study not showing this effect (Derrick *et al.*, 2016) is

[a)]Electronic mail: donald.derrick@canterbury.ac.nz the use of a two-alternative forced-choice (2AFC) paradigm in the former, and the use of an open-choice paradigm in the latter.

Open-choice responses allow individuals to independently produce a response to a question by saying it aloud or by writing it down. In contrast, closed-choice responses give preselected response choices out of which individuals are asked to choose the best (or correct) alternative. Moreover, closed-choice tasks deliver cues that may not be spontaneously considered, while in open-choice responses, the participants decide what information is relevant (Cassels and Birch, 2014). A task that exhibits demand characteristics or experimenter expectations (e.g., "Did the stimulus sound like 'pa'?") might give different results than one that does not (e.g., "What did the stimulus sound like?") (Orne, 1962). Closed-choice tasks are quite often influenced by demand characteristics. However, non-directed questions can be used to avoid this influence in open-choice tasks.

Closed-choice paradigms also present the correct answer as one of the options, whereas open-choice paradigms do not. The result is that people are more likely to focus on the correct answer with the former. Massaro (1998) found that speech stimuli were more frequently identified correctly in closed-choice tasks than in open-choice tasks. Colin *et al.* (2005) also found a higher percentage of the classic McGurk fusion response (e.g., perceiving auditory /ba/ and visual /ga/ as a fused /da/) for closed-choice tasks, at 19%–40% depending on audio intensity. Open-choice responses were more diverse, showing evidence of imperfect blends, and only at 0%–18% fusions. Mallick *et al.* (2015) replicated these findings when they examined the McGurk effect by modifying parameters like population, stimuli, time, and response type. They demonstrated that the frequency of the McGurk effect can be significantly altered by response type manipulation, with closed-choice response increasing the frequency of reported McGurk perception by 18% approximately when compared with open-choice responses for identical stimuli.

Open-choice paradigms therefore appear to allow perceivers more flexibility of interpreting multisensory stimuli, making their choices more telling of their experiences. However, the given answers may also be less accurate in relation to the underlying stimuli. The response choice may therefore be a significant contributor to variability in speech perception studies. As the 2AFC (Gick and Derrick, 2009) and the continuous speech open-choice study (Derrick *et al.*, 2016) are situated at both ends of the continuum of response-type paradigms, we currently do not know which response-type characteristic—if any—facilitates integration of audio-tactile stimuli during speech perception. It is therefore necessary to systematically study different response type paradigms situated along the 2AFC syllable identification–open-choice sentence identification continuum. One way to restrict the freedom of choice in an open-choice paradigm is by using a more restrictive stimulus type. In the continuous speech open-choice study (Derrick *et al.*, 2016), the participant had to identify five words in each sentence. Therefore, the stimuli used in the open-choice paradigm also differed from the syllabic and monosyllabic word stimuli used in the 2AFC experiments described above.

B. Stimulus type differences

Continuous speech stimuli like phrases or sentences are complex stimuli because they contain more information than is present in syllable identification tasks (semantic information, context information, utterance length, etc.). This additional information undergoes complex and higher order language processing. Sumby and Pollack (1954) demonstrated that individual open-choice syllable identification was much more difficult than two-word (tri-syllable) combinations, likely because the two-word phrase context helped provide more identifying information. Syllables are a simpler unit of language whose recognition does not require the same level of complex language processing, but also one that provides less contextual information.

Increasing stimulus complexity adds information to the system, aiding in speech perception accuracy, but making the cause of that increased accuracy much harder to identify. This is a serious issue in regard to the production of artificial air flow—complex audio-tactile speech generates complex air flow patterns. In a study contrasting voiced (/ba/) and voiceless (/pa/) plosives, the air flow difference can be simulated reasonably well by simply providing a short (50–80 ms) air flow for /pa/, and none at all for /ba/. With continuous speech, the air flow has to be time varying in a manner that is appropriate to the underlying complex speech. For the continuous speech perception experiment discussed above, we were able to do that reasonably well through careful recording of speech air flow outside the mouth, combined with careful electromechanical control of a dynamically varying air flow production system (Derrick *et al.*, 2015). However, we could not match speech air flow volume from speech, achieving only 1/12th that volume due to system constraint. These differences between speech air flow and simulated air flow did not matter for 2AFC experiments with single syllables, but may matter for more complex speech. Therefore, it is beneficial to reduce speech complexity in open-choice experiments to that of the 2AFC experiments in order to gain a detailed understanding of the factors that influence audio-tactile integration during speech perception.

When looking at the stimulus characteristics of syllables used in 2AFC experiments, place and manner of articulation have been shown to play a role. Derrick and Gick (2013), Gick and Derrick (2009), and Gick *et al.* (2010) show a consistent bias towards voiced labial and voiced alveolar identification. For example, when asked to identify whether they perceived an auditory-only /pa/ or /ba/ in noise, participants choose /ba/ more often. However, based on other literature, we do not expect similar results with an open-choice experiment. Lisker and Abramson (1964) identified the acoustic cue for perception of voicing and named it *voice onset time* (VOT), a measure of the air flow duration after stop release and before vocalic voicing. VOT is longer in voiceless velars than in labials (Lisker and Abramson, 1966). As a result, perceivers expect longer frication in voiceless velars than in voiceless labials (Zlatin, 1974). This makes a voiced bias in velars more likely. In addition, Benkí (2001) studied English stops by varying their formant transition duration, $F1$ frequency, and VOT. He concluded that bilabial and alveolar plosives are more likely to be perceived as voiceless than velar plosives. He also concluded that "Increasing $F1$ onset frequency and shortening transition duration also made voiceless judgments more likely" (Benkí, 2001, p. 1). However, the effect did not interact with place of articulation— bilabial and alveolar plosives were more likely to be perceived as voiceless no matter the manipulation (to both). Thus, based on the Benkí (2001) results, we would also expect voiceless bias for labials, and a voiced bias for velars. Taken together, these results suggest that we can expect a voiceless labial and voiced velar bias both for audio-only and audio-tactile conditions. However, given that we have always seen the exact opposite with labial onset (/ba/ vs /pa/) (Derrick and Gick, 2013; Gick and Derrick, 2009; Gick *et al.*, 2010) and alveolar onset (/da/ vs /ta/) (Gick and Derrick, 2009) syllable identification during 2AFC audio-tactile experiments, it is possible we will not see a voiceless bias for bilabials during this open-choice experiment.

C. Hypotheses

As noted, it is evident that the continuous audio-tactile study (Derrick *et al.*, 2016) has two major methodological differences from the 2AFC studies that preceded it: The first is the use of open- instead closed-choice, and the second is examination of sentence rather than syllable comprehension. The current study aims to differentiate the influence of both aspects by investigating the effect of response on its own, thus eliminating the sentence component. To achieve this, perceivers are asked to identify which syllable they perceive both for audio-only and audio-tactile conditions but using an open-choice paradigm where they type what they perceived rather than being forced to choose between two options. This will allow us to study whether forcing participants to choose between two given options, and thus restructuring the perception options, has been instrumental in studies showing successful audio-tactile integration. In addition, we sought to concurrently identify the interaction effects of place (labial vs velar) and manner (voiced vs voiceless) during open-choice syllable identification. This resulted in the selection of /ba/, /pa/, /ga/, and /ka/ as stimuli for the open-choice experiment. The study aimed to test the following hypotheses:

(1) Perceivers benefit from congruent auditory and tactile stimuli such that they more easily understand speech syllables when the two signals match real-world speech. The prediction is that perceivers will be able to identify syllables accurately 80% of the time in noisier conditions (lower signal-to-noise ratio) when the auditory and air flow (or absence of air flow) are appropriately congruent and available to the perceivers. That is, when there is air flow for /pa/ or /ka/ directed at the supra-sternal notch, but no air flow for /ba/ or /ga/. Because this study involves signal-to-noise threshold identification methods that differ from previous 2AFC experiments, the open-choice experiment results will be compared to those of a closed-choice one using the same threshold identification methods.

(2) There will be a significant influence of place of articulation on perception of manner such that perceivers are biased towards identifying velars as voiced and biased towards identifying labials as voiceless. The prediction is that the 80% accuracy signal-to-noise ratios (SNRs) for open-choice syllable identification will be lower for voiceless rather than voiced labial stimuli, and the 80% accuracy SNRs will be lower for voiced velar stimuli rather than voiceless ones.

II. METHODS

The methods of our study are divided into two sections, one for Experiment 1, an open-choice experiment, and one for Experiment 2, a closed-choice experiment used to compare and identify any differences in behavioural response during closed and open-choice experiments.

A. Experiment 1: Open-choice

1. Participants

For the experiment, 44 healthy participants were recruited. The University of Canterbury Human Ethics Committee reviewed and approved this study, and participants provided informed consent.

Participants then completed a demographic information sheet, reporting age, native language and history of speech, language and hearing difficulties. As part of the protocol, participants underwent an audiological screening. Pure tone audiometry testing was carried out for frequencies of 500 Hz, and 1, 2, and 4 kHz using an Interacoustics AS608 screening audiometer. Average pure tone thresholds were calculated and if the threshold was less than or equal to 25 dB hearing loss (HL), hearing sensitivity was considered to be within normal range. Of the 44 participants, ten participants did not meet language or hearing test requirements, and were excluded from this analysis, leaving 34 participants (30 females and 4 males) with a mean age of 23.2 years [standard deviation (SD) ± 6.4 years].

2. Materials: Auditory stimuli

Like Gick and Derrick (2009), this study uses four syllables: Two of which were identical to those used in the Gick and Derrick study (/pa/ and /ba/), having labial onsets, and two others (/ka/ and /ga/) had velar onsets. This one difference was intended to allow the kinds of fusion-based responses that occur in the McGurk effect (McGurk and MacDonald, 1976) described above, and also because the literature on place and manner interaction contrasts labial and velar places of articulation.

This study also contained the syllables /ka/ and /ga/ because we did not want participants to immediately guess that there were only two underlying choices - by interleaving four possible syllables, we restrict the likely choices without making the study into a de-facto closed-choice study. Also, previous literature suggested that articulatory features like place (as well as manner) of articulation were beneficial cues that contribute to perception of syllables (Eimas *et al.*, 1978; Lisker and Abramson, 1970; Miller and Eimas, 1977; Sawusch and Pisoni, 1974).

This study builds on the methodology of Gick and Derrick (2009), coupling an acoustic speech signal with small puffs of air on the skin. To produce stimuli for the experiments, we recorded a female native New Zealand English speaker in her mid 20 s with no speech or hearing disorders. She was recorded producing labial onset (/pa/ and /ba/) and velar onset (/ka/ and /ga/) syllables. The speaker produced twenty repetitions of each stimulus, as presented in randomized order on a computer screen placed in front of her.

The speech stimuli for this experiment were recorded in a sound-attenuated room using a Sennheiser MKH-416 microphone attached to a Sound Devices USBPre 2 pre-amplifier connected to a late 2013 15 in. MacBook Pro via USB cable. Recordings were done in Audacity to a 48 000 Hz pulse code modulation (PCM).wav file.

Syllables were matched for duration (390–450 ms each), fundamental frequency (falling pitch from 90 to 70 Hz) and intensity [70 dB(A)]. Using an automated process written in R (R Development Core Team, 2018), the speech token recordings were randomly superimposed 10 000 times within a 10 s looped sound file to generate an audio file containing speech noise for the speaker. According to Jansen *et al.* (2010) and Smits *et al.* (2004), this method of noise generation results in a noise spectrum virtually identical to the long-term spectrum of the speech tokens of the speaker and thus ensures accurate SNRs for each speaker and token. Speech tokens and the noise samples were adjusted to the same A-weighted sound level prior to mixing at different SNRs.

Recordings of the four underlying tokens of /pa/, /ba/, /ka/, and /ga/ were overlaid with this speech-based noise, generated using R (R Development Core Team, 2018) and FFMPEG (FFmpeg Developers, 2016). The SNR of the stimuli ranged from −20 to 10 SNR with 0.1 SNR increments. From −20 to 0 SNR, the volume of the signal was decreased, and volume of the noise was kept stable. From 0 to 10 SNR, the signal was kept the same volume and noise was decreased. Thus, the overall amplitude was maintained stable throughout the experiment.

3. Materials: Tactile stimuli

Aero-tactile stimuli were controlled by using an 80 ms long 12 kHz sine wave, aligned to the consonantal burst onset. The auditory signal was placed in the left channel, and the pump control signal for audio-tactile stimuli (and empty signal for audio-only stimuli) was placed in the right channel of a stereo audio file. The stored audio was used to drive a conversion unit that split the audio into a headphone out (to both ears) and the right channel to the air flow pump, mounted on the tripod, to release an air puff to the participant's suprasternal notch (at the base of the front of the neck) for the audio-tactile stimuli.

This speech air flow generation system used a Murata MZB1001T02 piezoelectric device (Tokyo, Japan) controlled through the Aerotak system (Derrick and De Rybel, 2015), as described in Derrick *et al.* (2014a). The Aerotak system uses the air flow signal to activate the Murata pump, which delivers air flow to the skin of the participant.

The pump has a 5%–95% rise time of under 10 milliseconds (contra the inaccurate 30 ms reported in Derrick *et al.*, 2015), with a maximum pressure of 1.5 kPa (15.29 cm H_2O, where normal conversational speech pressure caps at 7 cm H_20) at the source, and a maximum flow rate of 0.8 l/m, which corresponds to about one-twelfth of that of actual speech (normal speech volume around 11.1 l/m).

4. Procedure

Once the initial screening protocol was completed, participants were told that they might experience some noise and unexpected puffs of air along with syllables, consisting of a consonant and a vowel, during the task. Participants were asked to type the syllables that they heard on a keyboard for the computer in front of them and push the enter key to record their responses. Participants were then seated in a sound-attenuated booth and presented with the auditory stimuli via Panasonic RP-HT265 closed stereo headphones at a comfortable loudness level (approximately 70 dB). Aero-tactile stimuli were delivered to the suprasternal notch via the piezoelectric pump described above, positioned towards the subject's neck fixed at approximate 2.2 cm from the skin surface. The researcher stayed inside the experiment room with the participant during the experiment to make sure that pump placement was not disturbed and to ensure that participants were comfortable.

The 80% accuracy SNR was then identified using a fast and stable adaptive staircase method: The QUEST staircase (Watson, 1983). This method uses Bayesian estimation to place each trial at the most current probable SNR for the desired accuracy, reducing variability to sufficiently low amounts within 32 to 40 runs. As noted in the seminal paper on the topic, the method works because human psychometric functions are largely invariant when described using the log intensity (Watson, 1983). This method provides an efficient and therefore fast method of reaching an accurate SNR measurement.

Eight QUEST staircases, presenting audio-only and audio-tactile English syllables (/pa/, /ba/, /ga/ and /ka/), with 32 tokens each (consisting of four unique underlying recordings, each repeated eight times), were randomly presented to the participants through an experiment designed in PsychoPy software (Pierce, 2007, 2009) on a 2016 MacBook Air laptop.

The staircases were tuned to identify the 80% accuracy point. Participant accuracy was tracked by having the participants type out the perceived syllable into the experiment control program. Correct responses (typed "pa," "ba," "ga," or "ka/ca" based on the underlying auditory signal) were used to lower the SNR, and incorrect ones to increase it in order to obtain the 80% accuracy threshold. The length of time taken to identify one token was 6.5 s on average, resulting in 30-min experiments.

B. Experiment 2: Closed-choice

Experiment 2 focused on the audio-tactile component of a 2AFC audio-visual-tactile experiment, which itself has been preliminarily presented at a conference (Derrick *et al.*, 2018); completed results are currently under review. The full experiment is similar to experiment 1, except that it involved

12 staircases: eight QUEST staircases with congruent and incongruent visual stimuli, as well as four QUEST staircases with masked visual stimuli. Here we focus on the results from the four masked visual stimuli staircases as these allow direct comparison between audio-only and audio-tactile stimuli.

1. Participants

Ethics, recruitment, demographic information, and hearing tests were identical to those used in experiment 1. Participants included forty (40) New Zealand English perceivers, 18–46 years old ($\mu = 24.6$, SD = 8.0), 7 males, 33 female.

2. Materials

One female speaker, producing forty tokens of /pa/ [p^ha] and /ga/ [ka] each, was recorded in a sound-attenuated room with a professional lighting setup. Video was recorded on a Sony MediaPro PMW-EX3 video camera set to record with the MPEG2 HD35 hearing level (HL) codec, with a resolution of 1920 by 1080 pixels (16:9 aspect ratio), a frame rate of 25 frames per second (fps), and a hardware-synched linear pulse-code-modulation (LPCM) 16-bit stereo audio recording. The video was then converted to a time-preserving H.264 codec in yuv420p format encapsulated in an MP4 package, with audio extracted using FFMPEG (FFmpeg Developers, 2016). The audio was segmented in Praat (Boersma and Weenink, 2019), and the authors jointly selected ten recordings of each syllable that matched in duration, intensity, fundamental frequency, and phonation. In addition, the facial motion of each token was inspected to eliminate any case of eye-blink or noticeably distinguishable head motion.

3. Creation of A, AV, AT, and AVT stimuli

The ten /pa/ and ten /ga/ tokens were sorted by length to form the closest duration-matched pairs. Software was written in R (R Development Core Team, 2018), WarbleR (Araya-Salas and Smith-Vidaurre, 2017), FFMPEG (FFmpeg Developers, 2016), and the Macintosh borne-again shell (BASH). The software took the timing of each video file and extracted the video with 750 milliseconds lead time, and 500 milliseconds follow time. For each video stimuli, it produced a version with right-channel audio from the original and left-channel audio that was either empty (for no air flow stimuli), or contained an 80 millisecond 12 kHz maximum intensity sine-wave used to operate our custom air flow system. Twelve types of stimuli were produced to record audio-visual congruent, audio-visual incongruent, and audio-only. Here, we focus on the four audio-only (audio-still-video) stimuli where for each video, a version was produced with a blurred and still lower face. These four conditions include the congruent /pa/ with air flow and /ga/ with no air flow tokens, and the incongruent /pa/ with no air flow and /ga/ with no air flow tokens.

Speech noise, generated using the same procedure as experiment 1, was then overlayed on the right channel audio, making a video file for each token with signal-to-noise ratios from −30 dB to +15 dB, at 0.1 dB increments. The noise overlay was attenuated for all tokens above 0 dB, and the underlying audio was attenuated for tokens below 0 dB, ensuring that each token was of similar maximum amplitude for maximum comfort during the experiments.

4. Procedure

Stimulus presentation was identical to experiment 1, except that participants were seated with a screen behind glass positioned 1 meter from themselves so that they could see video as well as hear audio and feel air flow. The experiment presented 12 conditions interleaved into QUEST staircases with 40 tokens each, or 480 tokens total, taking about 20 min.

In addition, unlike experiment 1, two-alternative forced-choice (2AFC) QUEST adaptive staircases (Watson, 1983) were written in Matlab (The MathWorks, Inc., 2014) using the Psych Toolbox 3 software tools (Brainard, 1997; Kleiner *et al.*, 2007; Pelli, 1997). The QUEST staircases were tuned to identify the 82% accuracy threshold, with 40 trials for each of the 12 randomly interleaved blocks. The QUEST staircases used the standard Weibull function steepness (3.5), standard granularity of 0.01 dB SNR, and a wide latitude for allowable standard deviation (20 dB SNR) as per the protocol recommended in the Psych Toolbox manual. After each run, the QUEST quantile results, rounded to the nearest 0.1 dB, were used for the selection of stimuli, with the QUEST mean result used for final analysis, as per the recommendation in (Pelli, 1987). Initial SNRs for each staircase were tuned from a pilot experiment of ten participants set up similarly to the one described here, but with poorer quality video (Derrick *et al.*, 2018). This was done to prevent initial jarring perceptual differences during the first few runs. The initial values were /pa/ with air flow began with −8 dB SNR, /pa/ without air flow began with 0 dB SNR, /ga/ with air flow began with −10 dB SNR, and /ga/ without began with −8 dB SNR.

C. Data analysis—Both experiments

Descriptive statistics were run on each experiment (open-choice and closed-choice). Where appropriate, they are presented as tables of the means and standard deviations, expressed in dB SNRs for all the data by place, manner, and audio-tactile congruency. Notched box-plots were used to visualize variation of SNR based on place of articulation and stimuli type (audio-only vs audio-tactile). Generalized linear mixed-effects models (GLMM) were run using R statistical software (R Development Core Team, 2018), testing the interaction between *manner* of articulation [voiceless (/pa/ and /ka/) vs voiced (/ba/ and /ga/) stops], *place* [labial (/pa/ and /ba/) vs velar (/ga/ and /ka/) stops], and *congruance*, whether artificial air puffs were congruent (present for /pa/ and /ka/, and absent for /ba/ and /ga/) or incongruent (absent for /pa/ and /ka/, and present for /ba/ and /ga/) with the underlying acoustic stimuli. Note that only /pa/ and /ga/are used in experiment 2, so information from both place and manner combined will be described using *syllable* (/pa/ vs /ga/). Model fitting was performed in a stepwise backwards iterative fashion for each experiment separately. Starting with the most complex model allowed by the data, models were back-fit along the Akaike information criterion (AIC), to measure quality of fit. This technique isolates the statistical model that provides the best fit for the data. GLMMs were then run on both experiments combined, with interaction between experiment paradigm and audio-tactile congruency added to the interaction set. The models were then back-fit and are presented at the end of the results section.

R-Markdown files containing descriptive statistics and the full process of backwards-iterative model fitting reported in this paper, along with the code used to run the open-choice experiment and all the tokens needed for the experiment, as well as the code used to generate the audio files for the closed-choice experiment and run the experiment itself are provided online.[1]

III. RESULTS

Results will be described first for the open-choice experiment in Sec. III A, focusing on the effect of audio-tactile congruency and the effect of place and manner of articulation. This will then be followed by a presentation of the results of audio-tactile influence in the closed-choice experiment in Sec. III B. Finally, a direct comparison between the data from the two experiments will be described in Sec. III C.

A. Experiment 1: Open-choice

While 96% of the participants' responses were limited to those four syllables used in the underlying data (/ba/, /da/, /ga/, and /ka/, with /ca/ treated as /ka/), participants also responded with 32 other unique answers ("a," "aa," "ag," "ah," "aka," "b," "bah," "ban," "bu," "caa," "cal," "fa," "g," "gan," "ha," "k," "kaa," "ke," "la," "ma," "mba," "na," "p," "pa," "pan," "pla," "ra," "ta," "va," "wa," "ya"). These alternative answers span all of the English consonants but show remarkable consistency in vowel identification. That is, the participants demonstrated that they almost always recognized the vocalic portion of the stimuli, and so recognized that they were hearing speech. However, the consonants were harder to identify. Making the responses appropriate for a de-facto constrained open-choice experiment.

Five of the 34 participants had some portions of the experiment that reached a ceiling of +10 dB SNR. Therefore participant 2's /ka/ and /ba/, participant 6's /pa/, participant 8's /ka/, participant 14's /ka/, and participant 37's /ga/ and /ba/ data had to be excluded for hitting this maximum of +10 dB. Descriptive statistics were run on the 80% SNRs of the remaining data.

1. Descriptive statistics

The results of the experiment, by condition, are presented in Table I and Fig. 1. The results show that congruent presence or absence of air puff resulted in largely unchanged or slightly higher SNRs compared to incongruent tactile stimuli. The results also show a bias trend towards easier identification of voiceless labial and voiced velar responses. These trends are illustrated in the notched boxplots shown in Fig. 1. Removal of staircases reaching ceiling effects leaves behind few outliers, but the notches reveal the 1.73–4.24 dB SDs, and illustrate this dataset's variance across participants.

TABLE I. Mean and standard deviation of SNRs in the audio-tactile and audio only condition for different syllables. A = Auditory only, AT = Audio-tactile, C = Congruent stimuli, IC = Incongruent stimuli, SD = Standard Deviation. All numbers represent SNRs in dB.

Place	Labials				Velars			
Manner	/ba/		/pa/		/ga/		/ka/	
Condition	A (C)	AT (IC)	A (IC)	AT (C)	A (C)	AT (IC)	A (IC)	AT (C)
Mean	−4.52	−5.42	−5.27	−5.28	−8.81	−8.35	−4.19	−3.41
SD	1.73	1.92	2.12	3.37	3.50	2.57	4.24	3.76

After backwards-iterative model-fitting, the best-fit model for the open-choice experiment is shown in Eq. (1)

$$SNR \sim (place * manner) + (1 + (place * manner))|participant). \qquad (1)$$

In this model, the SNR at 80% accuracy were compared to the fixed effects. These included: (1) place of articulation (labial vs velar), (2) manner of articulation (voiced vs voiceless), (3) the interaction of place and manner, and (4) the full-factorial random effect covering place, manner, and audio-tactile congruency by participant. Note that this final best-fit model does not include whether auditory and tactile stimuli were congruent or not. The results of the fixed effects for this model are shown in Table II.

The results of place and manner are visualized in the notched boxplot in Fig. 2. The results clearly show that participants were only slightly biased towards identifying voiceless labial /pa/, but were biased towards identifying voiced velar /ga/than voiceless velar /ka/, as evidenced by the much lower 80% accuracy SNRs for /ga/.

Summarized, significant main and interaction effects for place and manner of articulation were shown in the open-choice syllable identification task. However, the results of this open-choice experiment do not show support for the influence of tactile stimuli on the accuracy of auditory speech perception in noise. Next, we examine the closed-choice experiment.

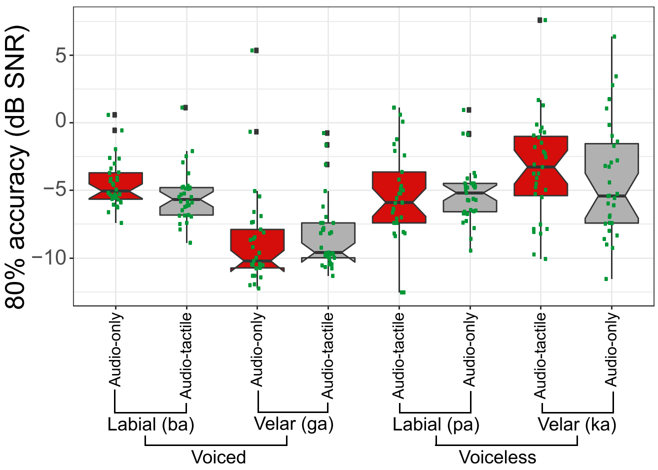

FIG. 1. (Color online) Notched boxplots demonstrating variability of SNR as a function of different stimulus conditions for experiment 1 (closed-choice). Red = congruent audio-tactile conditions. Please note that boxplots presented here center around median, while Table I lists the means. Gray = incongruent conditions.

TABLE II. Fixed-term results for the GLMM model shown in Eq. (1). *** = $p < 0.001$.

Fixed-effects	Estimate	Standard Error	*df*	*t*-value	*p*-value
Place (velar)	−3.57	0.504	79.4	−7.08	<0.001***
Manner (voiceless)	−0.335	0.511	84.5	−0.657	0.513
Place: manner	5.11	0.647	168	7.90	<0.001***

B. Experiment 2: Closed-choice

The results of the closed-choice experiment, by condition and token (/pa/ and /ga/), are presented in Table III. The results show that congruent presence or absence of air puff resulted lower SNRs than incongruent air puffs. This data stands in contrast to that seen for the open-choice experiment in Table I.

After backwards-iterative model-fitting, the best-fit model for the closed-choice experiment is shown in Eq. (2),

$$SNR \sim congruence + syllable + (1 + (congruence + syllable))|participant). \tag{2}$$

In this model, the SNRs at 80% accuracy were compared to the fixed effects of (1) audio-tactile congruence [congruent (/pa/ and air flow and /ga/ without air flow) vs incongruent (/pa/ without air flow and /ga/with air flow)], and (2) syllable (/pa/ vs /ga/), where syllable conflates place (labial /pa/ and velar /ga/) and manner (voiceless /pa/ and voiced /ga/). The results of the fixed effects for this model are shown in Table IV. These results appear quite different from the open-choice experiment, but these differences are more clearly illustrated through direct comparison of the open-choice and closed-choice experimental data.

C. Comparison of experiment 1 and experiment 2

The notched boxplots in Fig. 3, comparing the outcome of both the open-choice and closed-choice experiments show that the open-choice questions were considerably more difficult than the closed-choice questions, as evidenced by their much lower SNRs at 80% accuracy compared to the open-choice experiment. The boxplots also show that audio-tactile congruence has a greater effect on the notch position for the closed-choice experiment, especially for the labial data (/pa/). To allow direct comparison between the two experiments, only the data for underlying acoustical /pa/ and /ga/ are used from the open-choice experiment. As a result, the open-choice portion of the boxplots in Fig. 3 is not identical to the ones in Fig. 1.

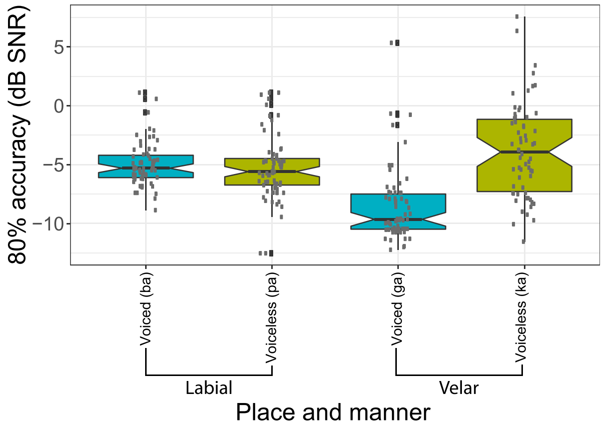

FIG. 2. (Color online) Notched Boxplots demonstrating variability of SNR as a function of manner (voiced and voiceless) and place (labial and velar) of syllable onset for experiment 1 (open-choice). Teal for voiced while gold is for unvoiced.

TABLE III. Mean and standard deviation of SNRs in the audio-tactile and audio-only condition for different syllables. A = Auditory only, AT = Audio–tactile, C = Congruent stimuli, IC = Incongruent stimuli, SD = Standard Deviation. All numbers represent SNRs in dB.

Manner	/pa/		/ga/	
Condition	A (C)	AT (IC)	A (IC)	AT (C)
Mean	−23.9	−20.0	−17.9	−16.0
SD	9.49	7.81	6.20	2.89

After backwards-iterative model-fitting, the best-fit model for the closed-choice experiment is shown in Eq. (3),

$$SNR \sim congruence * paradigm + (1 + (congruence + paradigm))|participant). \tag{3}$$

In this model, the SNRs at 80% accuracy were compared to the fixed effects of (1) audio-tactile congruence [congruent (/pa/ and air flow and /ga/ without air flow) vs incongruent (/pa/ without air flow and /ga/ with air flow)], (2) paradigm (open- vs closed-choice experiment), and (3) the interaction of audio-tactile congruence and experiment paradigm. Note that the random effects are quite simple—attempts to build more complex random effects resulted in the models failing to run. The results of the fixed effects for this model are shown in Table V.

The interaction plot in Fig. 4 shows the large difference in SNRs between the open- and closed-choice paradigms. The plot also shows that the audio-tactile congruent and incongruent results overlap for the open-choice experiment, but are separated by about 3 dB SNR for the closed-choice experiment.

IV. DISCUSSION

This study set out to test two hypotheses regarding open-choice syllable identification. The first hypothesis aimed to test whether tactile sensory information would congruently enhance and incongruently interfere with speech perception in an open-choice syllable identification task as it does in 2AFC experiments. The second hypothesis focused

TABLE IV. Fixed-term results for the GLMM model shown in Eq. (2). ** = $p < 0.01$, *** = $p < 0.001$.

Fixed-effects	Estimate	Std. Error	df	t-value	p-value
Congruence (incongruent)	2.85	1.04	41.0	2.73	0.009**
Syllable (ga)	5.03	1.17	41.0	4.28	<0.001***

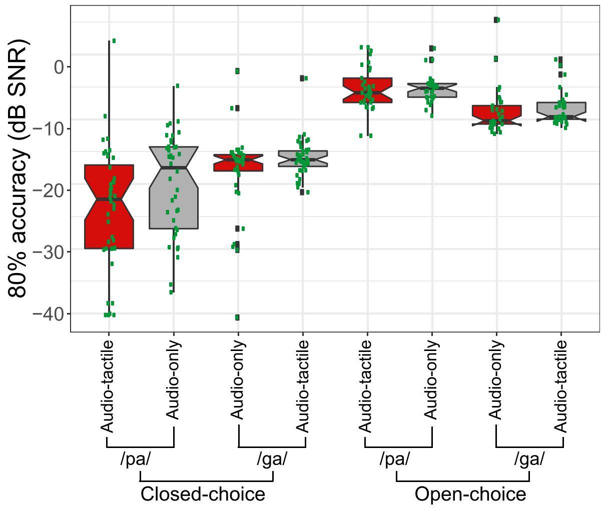

FIG. 3. (Color online) Notched boxplots comparing the 80% dB SNRs of the open- and close-choice experiments, with separate plots for syllable (/pa/ vs /ga/), and audio-tactile congruence. Red = congruent audio-tactile conditions. Gray = incongruent audio-tactile conditions.

on the influence of place and manner articulation on the threshold at which 80% correct syllable identification is reached.

The results of this experiment did not support the first hypothesis: Speech air flow had no discernible effect on syllable identification in an open-choice task. This lack of tactile influence on speech perception matches the results from a previous study of audio-tactile integration during an open-choice sentence identification task (Derrick *et al.*, 2016). The experiment presented here limited the choices to syllables, rather than words or sentences. In this way, this experiment disambiguated the potential influence of task-complexity from paradigm-induced choice-complexity.

The results of the open-choice experiment contrast clearly with the results of the similarly-designed closed-choice experiment. The best-fit model for experiment 1 showed no significant influence of audio-tactile congruence. In contrast, the best-fit model for experiment 2 showed enhancement in congruent audio-tactile conditions. Directly comparing the results of the two experiments showed a significant difference in the influence of audio-tactile congruence between the open-choice and closed-choice experiments, as seen in Table V, and made especially visually salient in Fig. 4.

These results therefore show that the response paradigm matters independently from syllable vs words-in-sentence identification. This open-choice task was, after all, considerably constrained in comparison to the task of identifying words in a sentence of continuous speech (Derrick *et al.*, 2016). After a few trials, participants in our open-choice experiment could possibly guess that they had to choose between four syllables, making this a constrained open-choice task. Yet despite this constraint, participants did sometimes type in many different answers, especially at the beginning of the experiment. Participants also demonstrated ambiguity of perception of the consonant as compared the vowel nucleus. This fits in with the observation that vowels are easier to identify in noise than stop consonants, with about double the accuracy in closed-choice identification tasks (64 choices provided) for any given SNR (Phatak and Allen, 2007).

TABLE V. Fixed-term results for the GLMM model shown in Eq. (3). * = $p < 0.05$, **, $p < 0.01$, *** = $p < 0.001$.

Fixed-effects	Estimate	Standard Error	*df*	*t*-value	*p*-value
Congruence (incongruent)	2.85	0	149	3.23	0.002**
Paradigm (open)	13.8	1.13	92.7	12.2	<0.001***
Incongruent: open	−2.61	1.31	150	−2.00	0.048*

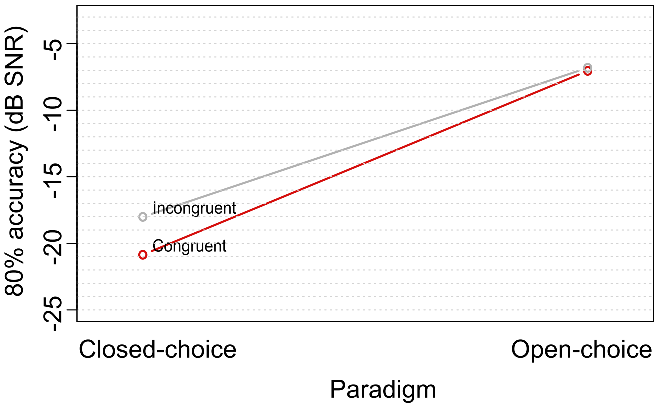

FIG. 4. (Color online) Interaction plot showing the relationship between paradigm and audio-tactile congruence.

Both the open-choice sentence-level (Derrick *et al.*, 2016) and the current open-choice syllable-level experiments failed to show an effect of audio-tactile congruency on speech perception accuracy. As the current study set out to examine the influence of response type on audio-tactile integration during speech perception, future studies should focus on distinguishing between the use of sentences rather than syllables on audio-tactile speech perception when response options are constrained.

The current air flow system was clearly suitable for identifying the influence of audio-tactile congruence in the closed-choice experiment 2; these closed-choice results provide good confidence that the same system was suitable for the open-choice experiment 1. However, as with the continuous speech perception research discussed in the introduction (Derrick *et al.*, 2016), the air flow system used for the present study had the same pressure (max 1.5 kPa) as speech but only one twelfth of the air flow (0.8 l/m) normally produced in the speech (11.1 l/m). This low air flow may have more strongly influenced the open-choice results because participants required clearer audio than for closed-choice syllable identification, resulting in a higher ratio of auditory to tactile signal intensity during the open-choice experiment. The low air flow rate from our pump system needs to be explored in future research, which is currently in the preparation states. This research should give us the basis for comparison needed to tell us if high air flow increases the benefit of the tactile component in speech perception.

The current state of aero-tactile speech perception research does not allow us to distinguish internal and external reasons for aero-tactile integration. We do not know if people are able to understand the connection of air flow and speech due to their motor control systems, or whether they learn the relationship as they learn to manipulate their glottis for speech. Two lines of research have been advanced currently to address these questions: (1) infant responses to speech airflow, and (2) research into directly measuring speech airflow outside the mouth in syllable, word, and phrase contexts using laser diffusion and pressure measurements. Following these measures, simulation of skin response based on what is already known from measuring skin response to braille (Phillips *et al.*, 1990) will allow us to know if the changes in air flow during more complex speech can affect the skin senses in a perceptually useful manner. In addition, we believe the origin of air flow could matter. Perceivers are probably less likely to experience air flow from /ka/ because /ka/'s plosive release is far inside the mouth, whereas /pa/'s plosive release is at the lips. This likely decreased experience with air flow from /ka/ as compared to /pa/ may have had an influence on the results. As a result, it would be interesting to study both the probabilities and travel distances of speech airflow for differing places of articulation. These studies will allow us to identify whether air flow can vary enough in real-world environments to allow perceivers to use such airflow to understand complex speech.

The second hypothesis aimed to test whether there would be a voiceless bias towards labial onset syllable perception and a voiced bias in velar onset syllable perception. The results of the open-choice experiment show support for hypothesis 2: Voiced velars were significantly easier to identify accurately than voiceless velars, and voiceless labials were easier to identify than voiced labials. It must also be emphasized that these results differed from previous 2AFC audio-tactile experiments (Derrick and Gick, 2013; Gick and Derrick, 2009; Gick *et al.*, 2010). For this open-choice experiment, the labial bias was towards easier identification of voiceless onsets. This result was, however, quite consistent with the results of Benkí (2001). The contrasting findings on the interaction between place and manner of articulation between the current open-choice study and previous closed-choice studies suggest that the processing of articulatory syllable information is also influenced by the response type paradigm. However, these results do not seem to interact with air flow effects in our open-choice experiment.

To conclude, the results of experiment 1, as compared to experiment 2, strongly support a distinction between two-way closed-choice classification and open-choice perception tasks. Stimulus characteristics related to place and manner of articulation have been shown to differentially facilitate speech perception accuracy in the open-choice study compared to previous closed-choice studies. In addition, air flow can affect behavioural closed-choice speech classification, but has not been shown to affect open-choice speech production. Taken together with previous results, this indicates that air flow may only produce statistically significant effects on speech perception in constrained closed-choice tasks.

ACKNOWLEDGMENTS

This research was funded by a New Zealand Ministry of Business, Innovation, and Employment "Smart Ideas" Phase II grant "Aero-tactile Enhancement of Speech Perception - Phase 2" to the first author. We would also like to thank the University of Canterbury for a Marsden Support Fund grant "How do our brains hear, see and feel speech?" to the first and third author.

[1] Tokens needed for the experiment, as well as the code used to generate the audio files for the closed-choice experiment and run the experiment itself are provided at osf.io/fy3qr/ and its supplementary materials.

Araya-Salas, M., and Smith-Vidaurre, G. (**2017**). "Warbler: An R package to streamline analysis of animal acoustic signals," Methods Ecol. Evol. **8**(2), 184–191.

Benkí, J. R. (**2001**). "Place of articulation and first formant transition pattern both affect perception of voicing in English," J. Phon. **29**, 1–22.

Bicevskis, K., Derrick, D., and Gick, B. (**2016**). "Visual-tactile integration in speech perception: Evidence for modality neutral speech primitives," J. Acoust. Soc. Am. **140**(5), 3531–3539.

Boersma, P., and Weenink, D. (**2019**). "Praat: Doing phonetics by computer (version 6.0.52) [computer program]," http://praat.org (Last viewed May 2, 2019).

Brainard, D. H. (**1997**). "The psychophysics toolbox," Spatial Vis. **10**, 433–436.

Cassels, T., and Birch, S. (**2014**). "Comparisons of an open-ended vs. forced-choice 'mind reading' task: Implications for measuring prespective-taking and emotion," PLoS ONE **9**(12), 1–20.

Colin, C., Radeu, M., and Deltenre, P. (**2005**). "Top-down and bottom-up modulation of audiovisual integration in speech," Eur. J. Cogn. Psychol. **17**(4), 541–560.

Derrick, D., and De Rybel, T. (**2015**). "System for audio analysis and perception enhancement," PCT Patent No. WO 2015/122785 A1.

Derrick, D., De Rybel, T., and Fiasson, R. (**2015**). "Recording and reproducing speech airflow outside the mouth," Can. Acoust. **43**(3), 102–103, available at https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2754.

Derrick, D., De Rybel, T., O'Beirne, G. A., and Hay, J. (**2014a**). "Listen with your skin: Aerotak speech perception enhancement system," in *Proceedings of the 15th Annual Conference of the International Speech Communication Association* (*INTERSPEECH* 2014), September 14–18, Singapore, pp. 1484–1485.

Derrick, D., and Gick, B. (**2013**). "Aerotactile integration from distal skin stimuli," Multisens. Res. **26**, 405–416.

Derrick, D., Hansmann, D., Haws, Z., and Theys, C. (**2018**). "Audio-Visual-Tactile integration in speech perception," in *LabPhon 16–The 16th Conference on Laboratory Phonology*, Lisbon, Portugal, p. 2.

Derrick, D., Heyne, M., O'Beirne, G., and Hay, J. (**2019**). "Aero-tactile integration in Mandarin," in *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, pp. 3508–3512

Derrick, D., O'Beirne, G. A., De Rybel, T., and Hay, J. (**2014b**). "Aero-tactile integration in fricatives: Converting audio to air flow information for speech perception enhancement," in *Proceedings of the 15th Annual Conference of the International Speech Communication Association* (*INTERSPEECH* 2014), September 14–18, Singapore, pp. 2580–2584.

Derrick, D., O'Beirne, G. A., De Rybel, T., Hay, J., and Fiasson, R. (**2016**). "Effects of aero-tactile stimuli on continuous speech perception," J. Acoust. Soc. Am. **140**(4), 3225.

Derrick, D., O'Beirne, G. A., Gordon, M., De Rybel, T., Fiasson, R., and Hay, J. (**2016**). "Effects of aero-tactile stimuli on continuous speech perception," in *5th Joint Meeting, Acoustical Society of America and Acoustical Society of Japan*, November 28–December 2, Honolulu, HI.

Eimas, P. D., Tartter, V. C., Miller, J. L., and Keuthen, N. J. (**1978**). "Asymmetric dependencies in processing phonetic features," Percept. Psychophys. **23**(1), 12–20.

FFmpeg Developers (**2016**). "FFmpeg tool [computer sofware]," http://ffmpeg.org/ (Last viewed May 2, 2019).

Fowler, C. A., and Dekle, D. J. (**1991**). "Listening with eye and hand: Crossmodal controbutions to speech perception," J. Exp. Psychol. Hum. Percept. Perform. **17**, 816–828.

Gick, B., and Derrick, D. (**2009**). "Aero-tactile integration in speech perception," Nature **462**, 502–504; Supp.

Gick, B., Ikegami, Y., and Derrick, D. (**2010**). "The temporal window of audio-tactile integration in speech perception," J. Acoust. Soc. Am. **128**(5), EL342–EL346.

Goldenberg, D., Tiede, M. K., and Whalen, D. H. (**2015**). "Aero-tactile influence on speech perception of voicing continua," in *Proceedings of the 18th International Congress of the Phonetic Sciences (ICPhS2015)*, August 10–14, Glasgow, UK.

Jansen, S., Luts, H., Wagener, K. C., Frachet, B., and Wouters, J. (**2010**). "The French digit triplet test: A hearing screening tool for speech intelligibility in noise," Int. J. Audiol. **49**(5), 378–387.

Kleiner, M., Brainard, D., and Pelli, D. (**2007**). "What's new in psychtoolbox-3?," in *Perception Thirtieth European Conference on Visual Perception Abstract Supplement*, 27–31 August 2007, Arezzo, Italy.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Lisker, L., and Abramson, A. S. (**1966**). "Some effects of context on voice onset time in English stops," Lang. Speech **10**, 1–28.

Lisker, L., and Abramson, A. (**1970**). "The voicing dimension: Some experiments in comparative phonetics," in *Proceedings of the 6th International Congress of Phonetic Sciences*, Academia Prague, Prague, Czech Republic.

Mallick, D. B., Magnotti, J. F., and Beauchamp, M. S. (**2015**). "Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type," Psychonom. Bull. Rev. **22**(5), 1299–1307.

Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioural Principle* (MIT Press, Cambridge, MA).

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

Miller, J. L., and Eimas, P. D. (**1977**). "Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions," J. Acoust. Soc. Am. **61**(3), 835–845.

Orne, M. T. (**1962**). "On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications," Am. Psychol. **17**(11), 776.

Pelli, D. G. (**1987**). "The ideal psychometric procedure," Invest. Opthalmol. Vis. Sci. **20**, 366.

Pelli, D. G. (**1997**). "The videotoolbox software for visual psychophysics: Transforming numbers into movies," Spacial Vision **10**, 437–442.

Phatak, S. A., and Allen, J. B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**(4), 2312–2326.

Phillips, J., Johansson, R., and Johnson, K. (**1990**). "Representation of braille characters in human nerve fibres," Exp. Brain Res. **81**, 589–592.

Pierce, J. W. (**2007**). "PsychoPy—Psychophysics software in Python," J. Neurosci. Methods **162**(1–2), 8–13.

Pierce, J. W. (**2009**). "Generating stimuli for neuroscience using PsychoPy," Front. Neuroinf. **2**, 10.

R Development Core Team (**2018**). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).

Sawusch, J. R., and Pisoni, D. B. (**1974**). "On the identification of place and voicing features in synthetic stop consonants," J. Phon. **2**(3), 181–194.

Smits, C., Kapteyn, T. S., and Houtgast, T. (**2004**). "Development and validation of an automatic speech-in-noise screening test by telephone," Int. J. Audiol. **43**(1), 15–28.

Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.

The MathWorks, Inc. (**2014**). *MATLAB and Statistics Toolbox Release 2014b* (The MathWorks, Inc., Natick, MA).

Treille, A., Cordeboeuf, A., Vilain, C., and Sato, M. (**2014**). "Haptic and visual information speed up the neural processing of auditory speech in live dyadic interactions," Neurophychologia **57**, 71–77.

Watson, A. B. (**1983**). "QUEST: A Bayesian adaptive psychometric method," Percept. Psychophys. **33**(2), 113–120.

Zlatin, M. A. (**1974**). "Voicing contrast: Perceptual and productive voice onset time characteristics of adults," J. Acoust. Soc. Am. **56**(3), 981–994.