RFID Tattoo: A Wireless Platform for Speech Recognition

JINGXIAN WANG, Carnegie Mellon University
CHENGFENG PAN, Carnegie Mellon University
HAOJIAN JIN, Carnegie Mellon University
VAIBHAV SINGH, Carnegie Mellon University
YASH JAIN, Indian Institute of Technology Bombay
JASON I. HONG, Carnegie Mellon University
CARMEL MAJIDI, Carnegie Mellon University
SWARUN KUMAR, Carnegie Mellon University

This paper presents an RF-based assistive technology for voice impairments (i.e., dysphonia), which occurs in an estimated 1% of the global population. We specifically focus on acquired voice disorders where users continue to be able to make facial and lip gestures associated with speech. Despite the rich literature on assistive technologies in this space, there remains a gap for a solution that neither requires external infrastructure in the environment, battery-powered sensors on skin or body-worn manual input devices.

We present RFTattoo, which to our knowledge is the first wireless speech recognition system for voice impairments using batteryless and flexible RFID tattoos. We design specialized wafer-thin tattoos attached around the user's face and easily hidden by makeup. We build models that process signal variations from these tattoos to a portable RFID reader to recognize various facial gestures corresponding to distinct classes of sounds. We then develop natural language processing models that infer meaningful words and sentences based on the observed series of gestures. A detailed user study with 10 users reveals 86% accuracy in reconstructing the top-100 words in the English language, even without the users making any sounds.

CCS Concepts: • **Human-centered computing** → **Accessibility technologies**; **Ubiquitous and mobile computing systems and tools**; • **Networks** → **Network services**;

Additional Key Words and Phrases: RFIDs, Battery-free Networks

ACM Reference Format:
Jingxian Wang, Chengfeng Pan, Haojian Jin, Vaibhav Singh, Yash Jain, Jason I. Hong, Carmel Majidi, and Swarun Kumar. 2019. RFID Tattoo: A Wireless Platform for Speech Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 155 (December 2019), 24 pages. https://doi.org/10.1145/3369812

1 INTRODUCTION

This paper seeks to develop an RF-based assistive technology for persons with voice impairments. In the US, more than 2 million people require digital Adaptive Alternative Communication (AAC) methods to help compensate for speech impairments [6]. While various classes of voice impairments exist, we target acquired conditions where users continue to be able to make facial and lip gestures associated with speech. We aim to learn these gestures

Author's address: Jingxian Wang, Chengfeng Pan, Haojian Jin, Vaibhav Singh, Yash Jain, Jason I. Hong, Carmel Majidi, Swarun Kumar Carnegie Mellon University.

2474-9567/2019/12-ART155 $15.00
https://doi.org/10.1145/3369812

over time to produce speech in real-time. Our approach applies to a wide range of temporary and permanent acquired dysphonia (voice disorders) ranging from hoarseness to complete loss of voice that occurs in about 1% of the global population [5].

While there is rich literature on assistive input-to-speech technologies for speech impairments, state-of-the-art solutions suffer important limitations. Camera-based [12, 14, 26, 33] visual solutions for real-time lip-reading require users to constantly be within line-of-sight of a camera, which may not be possible when the user is on the move. Audio-based assistive solutions [56] only apply to speech impairments where users are able to produce sounds and struggle in noisy environments. Past work has proposed a variety of face-worn sensors for speech sensing, particularly in clinical settings, such as magnets attached to the tongue [17], EEG helmets [66] and EMG electrodes on the face [38]. Assistive text-to-speech innovations require users to provide constant manual input to the system via keypads [3] or various user interfaces [9] that require training and practice for proficiency. There remains a gap for an everyday intuitive assistive technology for voice impairments that does not require external infrastructure, bulky sensors on the face or manual hand input.

We present RFTattoo, the first wireless speech recognition platform for voice impairments through skin-friendly, wafer-thin, battery-free and stretchable RFID tattoos. We fabricate specialized light RFID tattoos attached to the skin surface of face at known locations. Each tag is fabricated to be stretchable, flexible, wafer-thin, extremely light and made with hypoallergenic materials. The tags are designed to be hidden under makeup and extremely skin-friendly. We track the strain of individual tags over time as they deform in response to motions generated by different intended sounds. However, it is often the case that certain distinct sounds produce similar facial movements. To this end, we build natural language processing models that combine identified facial gestures in context to construct meaningful words and sentences. A detailed user study with 10 users reveals 86% accuracy in recognizing the top-100 words in the English language.

RFTattoo's first challenge is to process signals from RFID tattoos to recognize distinct facial and lip gestures called *visemes*[1] [27], that correspond to sounds the user intends to express. RFTattoo recognizes visemes by modeling the pure stretch of the flexible tag antenna. An intuitive approach to model tag stretch to infer its impact on the frequency at which it resonates. Specifically even a small change, say one millimeter, in the electrical length of an antenna lowers its resonant frequency by as much as 8 MHz in our experiments. Unfortunately, RFID tags in the U.S. operate in the FCC's unlicensed 900 MHz band with an effective bandwidth of 26 MHz. This makes it challenging to accurately capture the large frequency shifts induced by stretch. More importantly, requiring an RFID reader to hop through all frequencies even within the unlicensed band would be too time-consuming (~ few seconds) to recognize real-time speech.

RFTattoo addresses this challenge by probing multiple specially tuned RFID tags instead of probing multiple frequencies at the reader. In particular, we design an RFID tag that advertises the bits of its own current stretch value even if it is probed at one frequency (e.g. 915 MHz). Our approach to do so attaches multiple RFID chips to a common antenna, each tuned to multiple sets of specially chosen frequencies. We design the i^{th} chip in the RFID tag to respond at 915 MHz only if the i^{th} bit of the current stretch of the tag (expressed in millimeters) is one. In effect, this allows a tag to recover the bits of its own extent of stretch with a single frequency probe, with a small number of RFID chips per tag – logarithmic in the desired stretch resolution. We formulate a novel super-resolution optimization algorithm that improves this resolution even further by processing the power of the received signals across chips. Sec. 4 describes our approach and RFID tag design to retrieve stretch in greater detail.

A second challenge RFTattoo must address is the dynamic radio environment – changing orientation of the RFID tags, multipath reflections as well as movement of the user's body. RFTattoo achieves this through a novel tag antenna design that isolates the impact of stretch from other aspects pertaining to the radio environment.

[1] A viseme is a set of phonemes that look the same, for example when lip reading.

Specifically we fabricate two co-located RFID antennas with two materials – one stretchable and one non-stretchable. We then compare the signals received across both RFID tags to isolate any effect from the tag location, orientation and radio environment. Sec. 5 describes the materials and antenna design of RFTattoo tags.

Finally, RFTattoo builds a natural language processing framework to map stretch values of tattoos placed at different points in the face to recognize words and sentences the user intends to speak. A key challenge in this regard is the fact that some sounds produce identical facial and lip gestures (visemes) and therefore cause a high degree of ambiguity in the recognized phonemes. RFTattoo addresses this through two approaches. First, RFTattoo monitors subtle movements of the user's tongue through its impact on the magnitude and phase of the RFID tags on the skin's surface. We show how this allows for disambiguation of certain phonemes that produce identical facial movement. Second, RFTattoo leverages a useful property commonly exploited in natural language processing – the fact that adjacent phonemes are not completely independent but must follow the English dictionary and rules of grammar. Sec. 6 describes our approach to recognize common words and sentences at high accuracy, based on these observations.

Limitations: We emphasize a few important limitations of RFTattoo: (1) RFTattoo achieves highest accuracy when the location of RFID tags on the face are known through a light-weight calibration a priori. This means that for optimal performance, one must re-calibrate should RFTattoo tags be peeled off and on, or with natural wear. (2) RFTattoo may miss visemes should specific tags be unresponsive owing to shadowing from the body relative to the reader. (3) RFTattoo's accuracy is poor in the face of unknown or untrained words (e.g. less common words and proper nouns). This is a common problem shared by voice recognition systems [7] (e.g. Siri, Alexa, etc.) as well as visual lip reading systems [12, 14, 33]. We discuss and evaluate all these limitations in Sec. 8 and Sec. 9.

We implement RFTattoo by building custom tag antennas using stretchable Ag-PDMS conductors on PDMS substrates connected to three RFID chips. We use a meander-line antenna appropriately impedance tuned to respond at the 900 MHz ISM band. We use commodity Impinj RFID readers attached to the user's waist.[2] Our system is attached to the user's face using hypoallergenic stickers and covered with makeup. We conduct a detailed user study with 10 users including two users with temporary dysphonia (loss of voice). We also include results when all users are instructed to mouth words silently. Our results reveal that:

- RFTattoo achieves a median accuracy in stretch of 1.4 mm.
- RFTattoo distinguishes between eleven visemes of the English language [2] at an accuracy of 90%.
- RFTattoo recognizes the most frequently used 100 words of the English language at an accuracy of 86%.

Contributions: Our main contribution is a novel system that recognizes intended speech of users with voice impairments using light-weight RFID tattoos attached to the face. Our contributions include:

- Algorithms that recognize subtle mm-accurate stretches of the tattoos as well as movement of the tongue by processing RF-backscatter signals at a handheld reader.
- A natural language processing framework that recognizes various facial gestures associated with speech to construct meaningful words and sentences.
- A detailed user study that reveals the promise of our approach in recognizing intended speech, even when users do not make any sounds.

2 RELATED WORK

RFID-based on-body Sensing: Sensor-equipped RFID tags have been used to monitor temperature [71, 79], moisture [29], or even neural signals [78]. More recent work relies solely on the phase and RSSI of tags to

[2]Note while our implementation uses a relatively bulky 4-antenna RFID reader due to availability of channel state information, our system relies only on information from only one antenna. Our system will be compatible with much more portable and highly compact readers as future commercial products become more open.

accurately track their location [52, 77], including sensing the body skeleton [39], shape [40] and target imaging [72]. RFID tags have also been used for finger touch tracking by sensing the impedance mismatch between the tag chip and the antenna [62]. Recent work also utilizes coupling effect of the near-field antennas to distinguish different materials underlying RFID tags [32]. In contrast to these systems, RFTattoo seeks to infer the stretch of individual RFID tags at known locations on the user's face. It does so purely using signals received at a handheld RFID reader. We note that our system is designed to detect extremely subtle stretches (few mm) that are within the dimensions of an RFID tag from a handheld commodity single-antenna RFID reader. This is beyond the purview of state-of-the-art RFID location-tracking solutions and therefore necessitates new solutions.

Stretchable Electronics / Soft Robotics: There have been many recent advances in the use of soft materials for mechanically robust robots and electronics: crawling robots powered with compressed air [63], electrically-powered soft robots capable of locomotion through confined spaces [35], and self-healing robots that can continue to walk when their on-board circuitry is severed [55]. Advances in stretchable electronics have enabled soft circuits with multiple sensing modalities [34, 37, 49, 50], but these typically rely on battery-powered sensors or an external power supply. Stretchable RFIDs has been explored for strain monitoring [59, 68] and gesture detection [43, 46] but cannot be applied solely in the 900 MHz ISM band alone (per FCC limits) and are primarily based on RSSI and therefore vulnerable to multipath in the on-body context. In the HCI community, skin-friendly tattoos or epidermal electronics [41, 51, 54, 74–76] have been used to provide on-skin user interfaces, which either operate only in the near-field or require active components on the skin. In this paper, we focus on building passive, stretchable and flexible RFID tattoo that can sense speech with commercial RFID readers.

Automatic Speech Recognition: Recent research in automatic speech recognition for persons with voice impairments has explored visual and audio-based strategies [65]. Video-based systems such as lip reading by camera [12, 14] assume that the user is within the field-of-view of a camera. Audio-based solutions assume the user is able to make certain audible sounds and therefore do not apply to users with complete dysphonia (loss of voice). More recent work requires the use of specialized powered sensors to detect speech, for example: (1) detecting EMG signals when speaking by placing skin surface electrodes at the face [18, 42, 53]; (2) detecting tongue movements by tracking small magnets pasted on the tongue [17]; (3) detecting imagined speech by using EEG [20] or nerve signals [57]; (4) detecting air flow as a proxy for speech [28]; (5) detecting mouth movement of radio reflection using Wi-Fi [70]. These methods can work without any sound production, but either require intrusive equipment or are limited certain phonemes or words, precluding wide-scale deployment [80]. RFTattoo complements these approaches by providing a solution that detects a wide range of phonemes with a battery-free and light-weight RFID-based solution for speech recognition.

3 OVERVIEW OF RFTATTOO

RFTattoo's primary goal is to infer the speech of the user in real-time based on signals reflected off RFID tags. We assume the RFID tags are attached as light-weight tattoos on the skin at known locations on the face. We measure the signals reflected off these tags from a commodity portable RFID reader. We allow the reader to be portable and not in the line-of-sight of the tags. We measure the magnitude and phase of the signals reflected off multiple tags from the reader. We then process these signals to measure two physical properties: the stretch of individual RFID tags, and the position of the tongue. We monitor both these properties over time to classify between various visemes (facial gestures) the user makes. Finally, we process these visemes to infer the words and sentences spoken by the user. The rest of this paper describes three key components of RFTattoo (Fig. 1).

(1) Inferring RFID Stretch and Tongue Position: RFTattoo actively measures the stretch of each RFID tag and the position of the tongue – two key aspects that help recognize speech. RFTattoo infers stretch by monitoring its effect on impedance due to the fact that a stretched tag is longer and thinner. RFTattoo specifically measures the frequency response, i.e. the change in magnitude of the reflected signal across frequencies to study this effect.

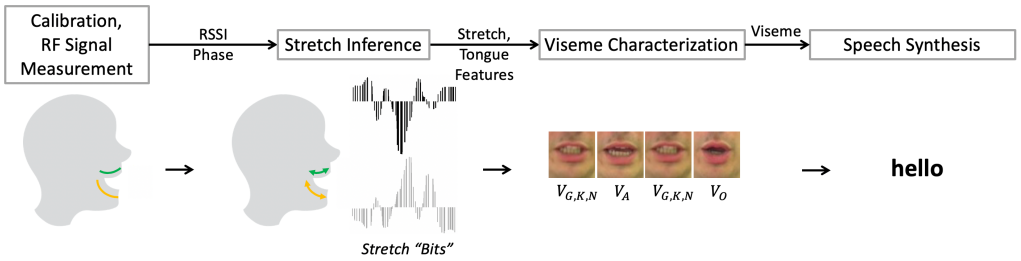

Fig. 1. RFTattoo's Architecture: (1) measures the wireless channel of RFID tattoos; (2) infers stretch bits and tongue position based on the reflected signal power and phase of multiple three-chip RFID tattoos; (3) feeds the features into machine learning models to recognize corresponding facial gestures (viseme – see Table.1); images show the corresponding viseme from the GRID dataset [23]; (4) combines visemes to form meaningful words and sentences by natural language processing.

We measure this property accurately and in real-time despite the limited bandwidth of commercial reader. We further show how RFTattoo can also infer the position of the tongue through its effect on RFID impedance. Sec. 4 details our solution.

(2) Designing the RFID tags: Next, we show how we can design RFID tags to specifically ease identification of stretch. We optimize both the material of the tag as well as the design of the antennas. First, the material of the tags must be optimized for facial skin – designing a material that is too thick will make it inconvenient to wear and designing one too thin will make it vulnerable to wear after repeated stretches. Our antenna design must also ensure that it resonates optimally with the RFID reader despite changing impedance when in contact to the skin and the limited area available on the skin. Sec. 5 describes our approach.

(3) Processing Speech: Finally, given the stretch of individual RFID tags, RFTattoo fuses these measurements to infer visemes, that are visual gestures of the face produced by different syllables pronounced by the user. We note that some visemes can be produced by multiple sounds, (e.g. "thee" and "tea" are indistinguishable visually). We show how we can disambiguate many such sounds using the position of the tongue. Sec. 6 describes our system that borrows from natural processing techniques to fuse the resulting phoneme measurements into meaningful words and sentences.

4 PROCESSING RFTATTOO SIGNALS

In this section, we characterize two properties from signals reflected off RFID tags on the user's face: the stretch of RFID tags and the position of the tongue.

4.1 Inferring Tag Stretch

Our key approach to monitor tag stretch measures the change in impedance as a result of the tattoo elongating. Specifically, as tattoos are stretched, its effecting width decreases and length increases, both of which increases its resistance and reactance. In effect, this causes a change in the resonant frequency of the RFID tag. For ease of exposition, we first assume the absence of multipath and a fixed relative orientation of RFID tag to the RFID reader. We will explicitly deal with these challenges later in the section.

Why does resonant frequency shift with stretch?: The resonant frequency of an antenna is the frequency, where the amplitude is higher than at adjacent frequencies. Stretching an RFID tag changes its antenna's electrical length and therefore its resonant frequency. Specifically, as the antenna length increases, the wavelength at which it resonates also increases meaning that the resonant frequency will shift towards lower frequencies.

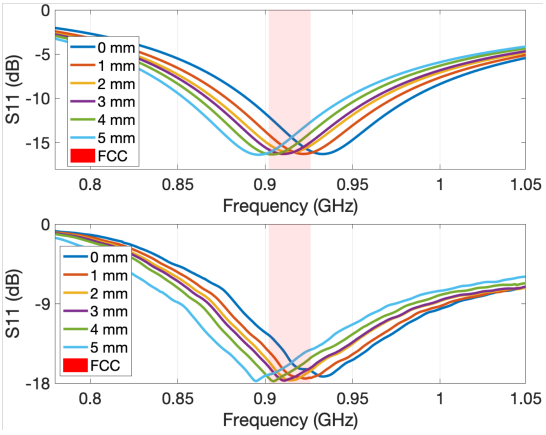

Fig. 2. Simulated data from HFSS (upper) and real data collected by VNA (down) of the resonant frequency shift with stretch steps of 1mm. Red patch shows the ISM bandwidth from 902 to 928 MHz. An average of 8 MHz frequency shift is seen. After just 3.25 mm of stretch, the resonant frequency goes out of the ISM band. Even within the band, the gain drops up to 6 dB.

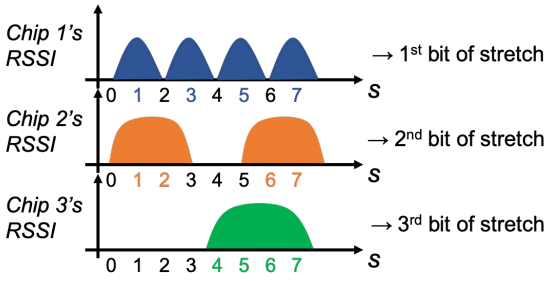

Fig. 3. The i^{th} chip on an RFTattoo tag responds with a maximum signal power when the i^{th} bit of s is one, where s is tag stretch. For example, Chip 3 responds when the 3^{rd} bit of stretch is on which corresponds to the possible stretch of 4, 5, 6, 7 in the figure.

Mathematically, the resonant frequency of a half-wave dipole antenna is written as [19]:

$$f_{res} = \frac{c}{2\sqrt{\varepsilon_e}} \frac{1}{L + L_e} \tag{1}$$

where L is the effective length of the half-wave dipole antenna and the L_e is the effective elongation of the antenna as stretching, ε_e is the effective relative permittivity of the antenna substrate and c is the speed of light in free-space. The ε_e can be estimated using the method mentioned from [36]. From this equation, we can see the resonant frequency is inversely proportional to its electric length leading to a very simple approach to infer stretch once electric length is found accurately.

Fig. 2 plots the frequency response of a half-wavelength dipole antenna when stretched to different lengths measured by the VNA and simulated by HFSS. Indeed it is surprising that a stretch of just 1 mm leads to a substantial resonant frequency shift of 8 MHz. On a positive note, this shows that resonant frequency allows for obtaining very fine grained values of stretch. However, on the flip side, detecting even an effecting stretch of a few millimeters will require an order-of-magnitude greater bandwidth when compared to the 26 MHz of bandwidth available in the 900 MHz ISM band. This motivates a challenging problem – how do identify resonant frequencies short of sweeping a wide range of frequencies.

How to find the resonant frequency?: A strawman approach to find the resonant frequency is to interpolate this by measuring RSSI across the available bandwidth at the 900 MHz ISM band (about 26 MHz). This constitutes a total of 50 hopping frequencies for Gen2 RFID protocol across 902 to 928 MHz (assuming the reactance is always conjugated with the RFID chip) accommodating a mere 3.25 mm of effective stretch in terms of resonant frequencies. Outside this range of stretch, we note that one can interpolate the RSSI drop across this small range of frequencies to infer the resonant frequency outside the FCC band. However, this approach remains vulnerable to multipath and noise. Further, it is time-consuming taking at least 10 seconds to sweep the 50 available frequencies with state-of-the-art readers [1].

RFTattoo finds the resonant frequency by relying on multiple co-located tags instead of multiple widely separated frequencies to obtain stretch. Specifically, we designing customized tags that all resonate at the 900 ISM band, but only when pulled to specific values of stretch. This means that by learning the distribution of signal power from the various tags, one can learn the precise distribution of signal power across frequencies.

Mathematically, let $t_1, \ldots, t_n$ be designed to resonate at a maximum power when stretched to values $s_1, \ldots, s_n$. Let $P(t_i)$ denote the power of the signal received from tag t_i. Then it is easy to see that the expected value of stretch s can be directly interpolated from the power of each tag as:

$$s = \sum_i s_i \frac{P(t_i)}{\sum_j P(t_j)}$$

The above approach has a unique advantage: it allows for finding the optimal value of stretch instantaneously without any frequency hopping beyond the FCC bands. However, it also has a notable disadvantage: it requires multiple co-located tags which can add to the bulk of the system. Indeed, using a larger number of tags can ensure more accurate system performance, given that it allows resonant frequencies to be more finely sampled. Indeed, it appears that the number of tags required is linear in the number of discrete resonant frequencies (i.e. stretch values) that need to be sampled. This leads us to a fundamental question: Can we design a system that requires a sub-linear number of tags $k << O(n)$ to sample n discrete resonant frequencies?

RFTattoo develops a solution that requires $k \sim O(\log n)$ tags to sample n discrete resonant frequencies. RFTattoo formulates the question of finding the resonant frequency through a divide-and-conquer approach. Specifically, let us suppose the optimal resonant frequency f can be represented in k bits ($k = \lceil \log n \rceil$). Then we design k tags each designed to always resonate on a set of multiple frequencies. In particular, the j^{th} tag is designed to resonate at all frequencies where the jth bit of the resonant frequency index f is one. In effect, this ensures that by simply reading off the power of the $k = \lceil \log n \rceil$ tags, one can infer the true stretch of all tags (see Fig. 3 for an example). Mathematically, we can write the index of the resonant frequency among n possible values as:

$$f = \sum_{i=1}^{k} 2^{i-1} x_i \tag{2}$$

where x_i is one if and only if $P(t_i)$ is above a threshold and zero otherwise.

How do we map resonant frequency to stretch?: In practice, resonant frequencies do not lie on discrete values but vary continuously over a large range of a few hundred megahertz. RFTattoo captures this by noting that a tag that resonates at a particular discrete frequency will also strongly reflect neighboring frequencies. Now let use $P(t_i, f)$ denote the power of the signal received from tag t_i at f frequency, notice that f is within the range of 902 to 928 MHz following the Gen2 RFID protocol. This means that we can effectively model the expected power $EP(t_i, f|s)$ by manually measuring the power of each tag $P(t_i, f)$ at specific frequencies should it be stretched to a value s. We model this expected power by simulating the frequency spectrum for each chip across a range of stretch values using HFSS [13] that draws from well-known antenna stretch models based on Poisson's effect [67] where Poisson's ratio is 0.5. For example, as depicted in Fig. 2, the orange curve shows the expected power for the tag with 1 mm stretch value under the FCC frequency spectrum. We observe that the simulated expected power across stretch values closely aligns with empirical measurements from our prototype (in Sec. 8). We compute the expected stretch as:

$$\tilde{s} = \min_s \prod_i |EP(t_i, f|s) - P(t_i, f)| \tag{3}$$

We then solve this optimization problem using standard sequential least square programming with a zero stretch of the tattoo as the starting point. We also bound the bandwidth and stretch within the 26 MHz and 50 mm. To achieve the above design we need to fabricate tags that have the ability to resonate at multiple specific

frequencies. In Sec. 5, we show how RFTattoo achieves this by adding multiple parallel impedances all coupling to the same chip.

Achieving Super-Resolution: We note that the above optimization problem deliberately models signal power to achieve high stretch resolution. Specifically, by simply treating whether a chip resonates or not at a specific stretch value as a binary 0 or 1, one can only resolve stretch at up to 2^k discrete values where k is the number of chips available on the tag. In contrast, as we model the continuous expected power quantity per chip with much more fined-grained stretch values instead of 2^k stretch, one can achieve significantly higher resolution than 2^k. Our results in Sec. 8 show median stretch accuracy of about a millimeter with only three chips per tag.

Impact of Tag Location, Orientation and Multipath: Our discussion thus far does not account for the unknown location of the RFID reader as well as multipath reflections from various surfaces between the reader and tag, including the user's body. In particular, the received signal power from an RFID tag depends upon four properties: (1) The location of the reader relative to the tag; (2) Signal multipath, including any attenuation and reflections of the user's body; (3) The orientation of the tag; and (4) The stretch of the tag. Our goal is to therefore isolate any received power change due to stretch from all remaining properties.

Our approach to do so attaches an additional RFID chip to the tattoo that is co-located but has a built-in antenna that does not undergo stretch. Given that this antenna is at the same location as the stretchable tag, it shares any attenuation across frequency owing to location, orientation and the effect of multipath. Indeed, any change in signal power across frequencies between this tag and the stretchable tag can be attributed purely to the stretch of the latter. Sec. 5 describes how we can use different antenna materials to create both stretchable and non-stretchable co-located tags.

4.2 Finding Tongue Position

RFTattoo uses the position of the tongue to disambiguate certain extremely similar sounds that produce virtually identical facial expressions, and therefore identical stretch values.

Finding tongue position: At a high level, our approach to find tongue position relies on the fact that the tongue changes the near-field radio environment of the tags that imposes additional impedance to the tags antenna. When the tongue moves from upper jaw front to lower jaw front, even though the tags shares the same surface of skin, the interior structure of the face changes so that the underlying material impedance changes. This would cause additionally resonant frequency shift apart from the stretch of the tags. We note that since tongue position is effectively a component of multipath, we only use signal measurements from the stretchable tag and do not use measurements from the non-stretchable tag. We rely on machine learning models to decouple the effect of tongue position from other sources of signal multipath. Specifically, we classify tongue positions using Random Forest at normal resting position, upper jaw front, upper jaw back lower jaw front and lower jaw back, while keeping the same all other sources of facial movement (see Fig. 10(a)). In Sec. 8.2, we evaluate system performance with two other candidate classifiers.

Finding the collective spatial arrangements of the tags: Besides the position of the tongue itself, RFTattoo also processes information from the spatial arrangement of various tags on the face to gather information on facial expression. We rely on the rich literature on RFID tag localization to do so. Our work specifically relies on WiSh [40] which allows a single antenna RFID reader to track the spatial shape formed by multiple RFID tags as they vary in geometry. We note that while the spatial arrangement of tags are valuable, the value of stretch plays a much more crucial role in determining facial expressions during speech. This is because movements of the face during speech are subtle, meaning that one has to compute minute changes within the dimensions of a tag itself to accurately characterize them.

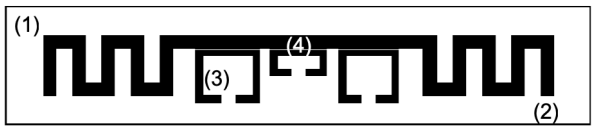

Fig. 4. RFTattoo's Antenna Design: (1) Stretchable substrate. (2) Half-wavelength Dipole. (3) Impedance matching networks with a sets of high impedance modules. (4) Coupled spacing.

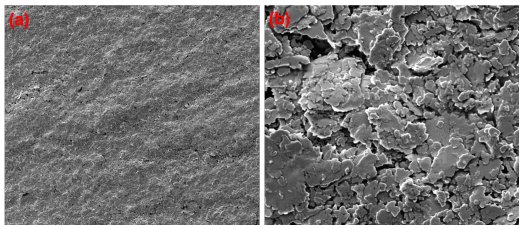

Fig. 5. Under stretched RFTattoo sampled from scanning electron microscope. (a) 300× magnified. (b) 5000× magnified. Small cracks appear between the silver flakes embedded withing the elastomer composite when the tag has been stretched.

5 TATTOO DESIGN AND FABRICATION

This section describes our RFID tattoo design and fabrication approach, specifically: (1) The design of the antenna pattern to fit within the dimensions of the user's face while providing high signal variations under stretch; (2) Materials and fabrication methods that maximize durability of the tags while keeping them light and thin.

5.1 Tattoo Antenna Design

Our objective is to design antennas that resonate at multiple frequencies and respond over well-defined bandwidths when stretched to a specific value. Doing so requires carefully designing the impedance of the RFID circuit. RFTattoo designs three components of the tag to achieve this: (1) the dipole, which is the physical far-field component that enables the communication between the tag and the RFID reader; (2) the inductor loop matches the impedance of the RFID chip; and (3) the coupling section, which transfers power between the dipole antenna and the inductor loop. We describe how each of these components are designed below:

Designing the dipole to be skin-compatible: RFTattoo uses a center-fed half-wavelength dipole antenna. The electrical length of a half-wave dipole is half-wavelength of the operating frequency of the antenna. However, at the 900 MHz ISM band, the physical length of a half-wave dipole is approximately 16 cm, which is too long for a face tattoo. To reduce the size of the antenna, we adapt the technique of meandering design (see Fig. 4) that maintains a reduced size with a long reading range [59]. We tune the number of meandering bends of the antenna in a manner that minimizes antenna loss (i.e. increase trace width), subject to length constraints of the antenna to resonate at the appropriate frequency.

As we design the dipole antenna, we need to consider the impact on the resonant frequency of the material to which the tag is attached – the human skin, in our case. Skin has heavy dielectric properties which motivates the need to appropriately tune the number of meander bends [60]. In our antenna design process, we build an ANSYS simulation that models the impact of human skin through a multiple substrate layer – containing one layer of antenna substrate and one layer of a human model provided by ANSYS HFSS. Our simulation is designed to optimize the performance of the dipole to resonate at 915 MHz. Having said this, in practical experiments, we observe minor variations in the resonant frequency between different users due to differences in the electric properties of their skin. Fortunately, the impact of this shift remains consistent across stretch of an RFID tag and can therefore be calibrated for a priori when the user makes a neutral expression. In Sec. 8, we demonstrate how our approach generalizes to diverse types of human skin.

Impedance Matching to resonate at many frequencies: An important challenge we face in antenna design is to make it respond to multiple distinct frequencies. One option is to pack together multiple antennas of different lengths co-located in the vicinity of an RFID tag. Doing so would clearly be too bulky to attach as a wafer-thin

tattoo. In contrast, RFTattoo tunes a single dipole antenna to multiple resonant frequencies through a multiple impedance matching units.

To elaborate, recall that to maximize the power transfer efficiency between the chip and the antenna at a given resonant frequency, any impedance mismatch should be resolved between these two components. A tag's reflection coefficient (Γ) which accounts for the impedance mismatch between the chip and the antenna is given by $\Gamma = \frac{Z_C - Z_A}{Z_C + Z_A}$, where Z_A is the impedance of the antenna and Z_C is the impedance of the chip. To maximize the efficiency of wireless power transfer to an RFID tag, antenna impedance is tuned to be conjugate matched with the input impedance of the chip at its center frequency, i.e. $Z_A = Z_C^*$. We note that a matching circuit is necessary even if an RFID tag is tuned to a single resonant frequency, as it is inevitable that some impedance mismatch remains between a dipole antenna and chip, no matter how carefully designed.

Notice that while impedance may mismatch between chip and antenna at one particular frequency, it is simultaneously possible that the impedance could perfectly match at a different frequency. We rely on this notion to provide multiple resonant frequencies for one chip. In particular, we build a multiple impedance matching network for each co-located chip. Our method uses a set of parallel inductively coupled loops which can also be seen as multiple high impedance coupling units for one chip to source the dipole [24]. Each loop is composed by series LC components and a resistor. Each branch of the series LC components couples with the stretchable dipole antenna (Fig. 6(a)). Further, each branch of the series LC components also matches the impedance with the feed terminal at different frequencies shown in Fig. 6(b). The advantage of such an approach is that the input reactance of the dipole and each loop now is only depends on the loop inductance at the corresponding required resonant frequency. Thus, we can control the reactance of each loop by carefully designing it to directly match the reactance of the chip at multiple discrete frequencies so that $X_{loop}(f) + X_A(s, f) = -X_C(f)$ where X_{loop}, X_A and X_C are the reactance of the loop, antenna and the chip respectively. For example, when dipole antenna is stretched at s_1, the sum reactance of dipole antenna and the loop (L_1, C_1) will match with the reactance of the chip at frequency f_1, while the other loops mismatch at this particular frequency.

Inductively Coupled Spacing to tune bandwidth: Besides the center frequency, RFTattoo also must tune the bandwidth over which a tag resonates. This is important since RFTattoo may need tags respond to a range of consecutive closely spaced frequencies. RFTattoo achieves this by tuning the position of the inductor loop [64]. Recall that this directly impacts the power transfer between the chip and the dipole via mutual coupling. Specifically, a smaller spacing between the loop and dipole will increase power transfer efficiency while ensuring a narrower band tag. RFTattoo therefore achieves varied bandwidths by adapting the loop-dipole spacing. We further do so differently for each chip to reserve different desired bandwidths as shown in Fig.3. Our evaluation in Sec. 7 uses a loop-dipole spacing of 1 mm, 0.6 mm, 0.01 mm respectively.

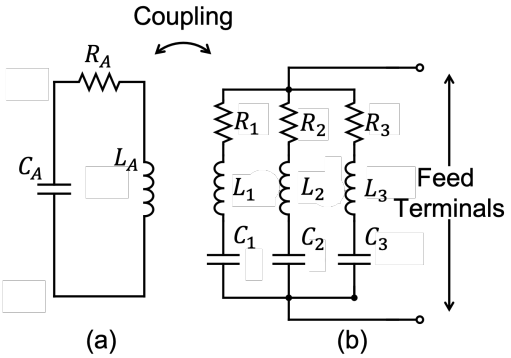

Fig. 6. (a) Equivalent circuit of dipole antenna. (b) Equivalent circuit of multiple impedance matching networks composed of parallel inductively coupled loops. Overall, this figure shows that multiple parallel coupled loops intend to match with the impedance of feed terminals (chip) at different frequencies which in turn leads the chip to resonate at many frequencies (three frequencies shown above).

5.2 Tattoo Material and Fabrication

This section describes our choice of materials in designing a stretchable RFID tag. A key trade-off that dictates our choice of materials is the balance between structural integrity and thickness. Thicker tags are more robust to repeated stretching and maintain integrity, yet are more uncomfortable to attach to skin. We therefore choose materials that provide the maximum structural integrity possible, while remaining as thin as 0.4 millimeters.

Material Selection of Stretchable Conductor Ink: The stretchable conductor is composed of a percolating network of electrically conductive filler embedded within a stretchable polymer matrix. Silver (Ag) flakes are selected as the filler due to their high conductivity (6.3×10^7 S/m) since this allows for stretchable traces with low electrical resistance. In addition, the platelet-like geometry of the flakes can reduce the electromechanical hysteresis of the composite during tensile loading and unloading cycles. We use poly(dimethylsiloxane) (PDMS) as the polymer matrix due to its low Young's modulus, high strain limit, and high elastic resilience (i.e. negligible mechanical hysteresis during loading and unloading). In order to achieve an adequate volumetric conductivity without sacrificing the elasticity of the matrix, we use a 30% volume fraction of Ag. This allows for a conductive elastomer composite with reasonable stretchability and conductivity (3000 S/cm). In addition, the printability of conductive ink is another key property for either fast prototyping or scalable manufacturing. To enable printability, we use methyl isobutyl ketone (MIBK) or order to reduce the viscosity of ink composite. The detailed fabrication process of the Ag-PDMS ink composite is presented in the Sec.7.

RFID Tags Fabrication: The RFID tags are compose of stretchable antennas, i.e. Ag-PDMS conductors, on a stretchable substrates. PDMS is used as the substrate to hold the unsolidified Ag-PDMS ink during curing and also to support the Ag-PDMS conductive traces during mechanical deformation. PDMS is particularly well-suited for this since it has an elastic modulus that is similar with human skin and is comfortable to wear. Moreover, the strong bond between PDMS substrates and Ag-PDMS composite facilitates the integrity of the RFID tags during mechanical deformation.

Stencil printing is used to pattern the Ag-PDMS and enables fast prototyping of the stretchable RFID tags. A commercially available Ag-filled conductive epoxy is used to solder the Ag-PDMS antennas with the packaged RFID chips. Typically, during stretching, there will be stress concentrations at the interface between the stretchable conductor and rigid RFID chip due to the compliance mismatch between the two. This stress concentration can lead to mechanical failure. To prevent this, we encapsulate the chip with an additional layer of PDMS in order to reduce the stress concentration and improve mechanical robustness.

Stretchable tag vs. Non-Stretchable tag: As mentioned in Sec. 4.1, RFTattoo uses a co-located stretchable and non-stretchable tag to account for location, orientation and signal multipath. While our approach above details the design of the stretchable tag, we also designed non-stretchable tags. We built the latter using copper-based antenna made from a commercially available flexible PCB (Pyralux FR8510R, DuPont) through solid ink (wax) based prototyping method [21]. In our experiments, we also use a miniature commercial RFID tag as the non-stretchable tag to be attached to the center of the stretchable tag.

6 FROM RF SIGNALS TO SPEECH

RFTattoo synthesizes speech by processing the stretch, location and tongue position of various points in the skin obtained from the RFID tag signals. It first uses this information to classify between various facial gestures called visemes that are unique to different sounds. RFTattoo then borrows from the rich literature on text-to-speech in natural language processing to synthesize speech in real-time.

Table 1. Phoneme to viseme clustering mapping [48]

Viseme Class	Phonemes Set	Sound
V_A	/ɑː/ /aʊ/ /aɪ/ /ʌ/	car, house, fly
V_E	/e/ /eɪ/ /æ/	egg, same, cat
V_I	/iː/ /ɪ/	sheep, ship
V_O	/ɔː/ /ɔɪ/ /əʊ/	door, coin, nose
V_U	/ʊ/ /uː/	book, boot
$V_{J,CH}$	/ʤ/ /ʧ/ /ʃ/ /ʒ/	jeep, cheap, ...
$V_{P,M,B}$	/p/ /b/ /m/	pit, bit, map
$V_{F,V}$	/f/ /v/	fat, vat
$V_{D,T,S}$	/d/ /t/ /s/ /z/ /θ/ /ð/	din, tin, ...
$V_{R,W}$	/r/ /w/	run, we
$V_{G,K,N}$	/g/ /k/ /n/ /ŋ/ /l/ /y/ /h/	gut, cut, ...

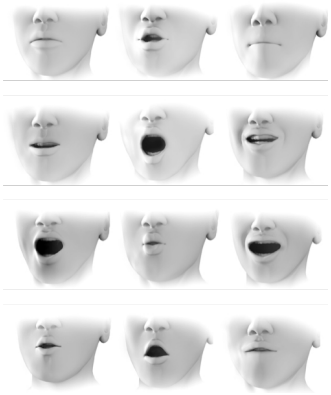

Fig. 7. Samples of visemes [10]. 12 visemes are used to map to all possible phonemes including the rest pose.

6.1 Characterizing Visemes

A *viseme* is a unit of visual speech – more specifically, the visual equivalent of a phoneme (a unit of sound in speech recognition). Each viseme represents the shape of the face when the user attempts to speak a particular phoneme [27]. Past work on automated lip-reading have widely used visemes to recognize speech based on video input[16]. Recognizing shorter visemes as opposed to longer speech segments has several advantages such as needing less training effort and generalizing well for different speaker identities (speaking styles, accents, etc.).

Choice of Visemes: Phonemes map many-to-one to visemes, because many phonemes can not be distinguished using only visual cues (e.g. "p" vs. "m" sounds). Phoneme-to-viseme mappings have been constructed mainly by two approaches: linguistic and data-driven. In this paper, we use the map from [48] obtained through a hybrid linguistic and data driven approach – a relatively sparse set which worked well experimentally (see Sec. 8). This map is composed by 38 phonemes and 11 classes (plus a silence class) in Table 1.

Viseme Classification: To classify the different viseme sets, we use the resonant frequency shift property of the RFID tattoo tag. RFID tattoos are attached to 4 different locations on the persons face: above the upper lip, below the lower lip, the left cheek and the right cheek. As the person utters the different phoneme sounds, these tags are stretched by different amounts resulting in diverse set of resonant frequencies. The response of all these tags are collected to obtain tag stretch as detailed in Sec. 4.

To train the classifier, we extract an extensive set of features mentioned in Table 2, based on the time and frequency properties of each tag's response. In particular, these features are extracted from tag stretch, tag and tongue location as well as normalized RSSI and phase values of each of the tags' response over time and frequency (within FCC bins). Notice that stretchable tags and reference tags have weak mutual coupling effect with each other. Small fluctuations in the RSSI value could occur, before extracting the features, we smooth the raw data over a 0.05 second time window. We select only the 125 most important features to train the classifier model after carrying out a set of relevance tests on the initial feature set. We train our classifier using 9 different classification models: k-Nearest Neighbors, RBF Support Vector Machine, Gaussian Process, Decision Tree, Random Forest, AdaBoost, Naive Bayes, Gaussian Mixture Model and QDA. We use Random Forest in classifying visemes which achieves the best accuracy (90%) among classifiers.

6.2 Speech Synthesis

Now we aim to synthesize the speech that the user intends to speak. At each unit of time, we predict a list of phoneme candidates, derived from the viseme mapping (see Table 1), as well as their likelihood scores.

The likelihood scores are obtained by our machine learning model, which outputs the predicted viseme with corresponding probabilities for all possible visemes. Using these phoneme candidates, we can reconstruct words with ambiguities. We note that despite the 90% accuracy in viseme classification and even upon accounting for tongue position, the ambiguity of the phonemes could significantly impact speech reconstruction accuracy.

To address this, RFTattoo draws from a salient advantage of natural language processing – adjacent phonemes and words are not independent – they are limited by the English dictionary and rules of grammar. We leverage this fact to disambiguate the recognition results and recognized the transcript of what the user intends to speak. Finally, we synthesize the speech using a public text-to-speech API[3].

During the operation, our recognition algorithm will produce a prediction stream that contains the recognition result and the corresponding time windows. Each recognition result r_k consists a list of phoneme candidates $r_k c_1, r_k c_2, ..$ and associated likelihoods $r_k l_1, r_k l_2, ...$

Word & Sentence Segmentation: To perform speech recognition, we first organize the recognized phonemes into words and sentences, based on their recorded time stamps. Here, a word is comprised of phonemes, and words compose a sentence. The lengths of pauses between in-word syllables, words, and sentences often vary. We run a pilot study with four participants and empirically determine the pause thresholds for both word-level separation and sentence level separation. If the pauses between multiple adjacent phonemes are smaller than the word/sentence threshold, we group these phonemes into the same word/sentence.

On-the-fly Word Disambiguation: Next, for each set of phonemes constituting a word, we derive a set of most likely word candidates using a pronouncing dictionary [30]. A pronouncing dictionary defines the mapping between sequences of phonemes and words. We then need to select one word from each group to assemble the final sentence. Choosing the words randomly, or even the most likely word per phoneme sequence often results in gibberish, since the words may not form meaningful sentences in combination. Leveraging this fact, we build a Bayesian model to evaluate the naturalness of the sentence formed by different word sequences. Let $N(w_k|w_1, w_2, ..w_{k-1})$ denote the naturalness score of choosing the word w_n among a group G_k, given that the prior sequence $w_1, w_2, ..w_{k-1}$ is determined. The selection of the incoming word W_k^* is equivalent to finding word that can maximize:

$$w_k^* = \arg\max_{w_k} N(w_k|w_1, w_2, ..w_{k-1}) * l(w_k) \tag{4}$$

where $l(w_k)$ is the word likelihood score from the earlier viseme recognition.

We use a co-occurrence relation to measure the naturalness. We count the frequencies of m consecutive words appear in a large document collection, and use these frequencies to indicate the naturalness. In other words, the more common the word sequence is, the more natural the sequence would be. We measure the co-occurrence in a sliding window of m consecutive words.

$$N(w_k|w_1, w_2, ..w_{k-1}) = \prod_{1}^{m} p(w_k, w_{k-1}, .., w_{k-m+1}) \tag{5}$$

where $p(w_k, w_{k-1}, .., w_{k-m+1})$ is the non-zero frequency of these consecutive words in a large document collection. If we cannot find a specific consecutive word sequence in the corpus, We set $p = 1e^{-3}$ to avoid multiplication by zero. Our implementation sets $m = 3$ and measures the frequency in Cornell Movie Dialog Corpus [25]. While our approach is simple and easy to reproduce, a more specific and contextual corpus, such as including sentences used most commonly in daily conversation can improve performance.

[3]https://cloud.google.com/text-to-speech/

7 IMPLEMENTATION

Hardware of RFID Tattoo: We build the dipole and inductor loop of RFID tattoo using an Ag-PDMS composite to make the antenna of tattoo maintain high conductivity as well as stretchablity. We use PDMS as the substrate of the tattoo, which provides up to 45 mm elongation for the selected tattoo dimensions. We use three LXMS31ACNA [8] chips to embed in each stretchable RFID tattoo (Fig. 8(b)). In our experiments, we use one Xerafy [11] chip as the non-stretchable reference tag which is placed on the center of each RFID tattoo. After careful fabrication and multiple interactions of adjustment, our wafer-thin RFID tattoos have a dimension of 80×10×0.5 mm.

Participants in the IRB Study: Consent forms[4] for participation in the research study were obtained at the time of the study, and all participants received a comprehensive description of the experimental procedures. We hold an active IRB protocol, which allows for attaching RFID tags to monitor bio-signals from human participants. We test our system with 10 participants from different genders, ages from 24 to 50 and include two participants who have temporary dysphonia. During the experiments, participants with healthy speech ability are also asked to speak without sound.

Evaluation and System Setup: We fully implement RFTattoo on an Impinj RFID reader attached to the user's waist. We note that we use a relatively bulky 4-antenna RFID reader as it provides wireless channel information and is readily programmable [40]. Our system however relies on information from only one antenna, allowing for much more portable and compact commercial readers in future commercial deployments. Our reader uses frequency-hopping spread spectrum to hop fifty frequencies across 902 MHz to 928 MHz (ISM band) using a pseudo-random sequence every 200 milliseconds. RFTattoo's algorithms are implemented in Python in real-time. The RFID reader is connected with only one single antenna and is carried by the participants. We test the participants in static and mobile indoor multipath-rich office space with walls, cubicles and furniture. We chose to attach four customized RFID tattoos on the upper and lower jaw and two sides of mouth which captures the most significance features of speech [44]. We mount the antenna with different relative distances to the users face resulting in three different levels of signal-to-noise ratio. Unless specified otherwise, all our experiments below use a multipath-rich indoor office space with wall and furniture, with the reader along the user's waist and often in non-line-of-sight relative to tattoos on the face.

Calibration Procedure: Unless specified otherwise, prior to the experiments, the participants are involved in doing a light-weight calibration where the relative location of RFID tags on the face are calculated and calibrated a priori. This means that the system is re-calibrated should RFTattoo tags be peeled off and on, or with natural wear. Also, we ensure that antennas of adjacent tattoos do not overlap with each other have a gap of at least 3 mm between them. Once participant attaches the four tags to their face, they are asked to rest in place and speak three words to test the signal level of the chip responses. Since the tags are specifically designed to operate on human skin, if the tags are not properly attached to the skin, the tag response level will be weakened and limit the communication distance of our system. We also calibrate for a priori when the user makes a neutral expression to mitigate minor variations in the resonant frequency between different users due to differences in the electric properties of their skin. The whole process requires less than 2 minutes. We report our accuracy for cases where per-user prior calibration or training was not performed in Sec. 8.4.

Data Management: We split all collected data into separate random training and testing data with the size ratio to be 4:1 [45, 47]. We evaluate a score of the training dataset using cross-validation over five folds for overfitting analysis. Then we train our model using the training dataset. Once we have our trained model, we predict and evaluate the accuracy of the testing dataset without using cross-validation. In the paper, all accuracy is testing accuracy.

[4]This work does not raise any ethical issues

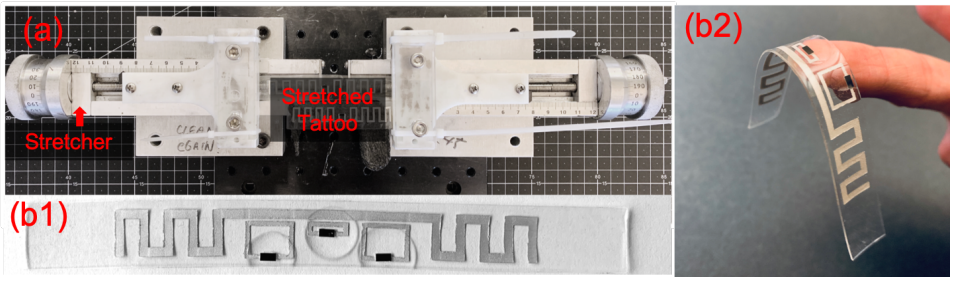

Fig. 8. (a) Rotary system for stretch evaluation of RFTattoo. (b) RFID Tattoo.

8 RESULTS

We first present our results from two microbenchmarks evaluating stretch and tongue position. We then evaluate the performance of our speech recognition system at three different levels: viseme, word and sentence.

8.1 Accuracy of Stretch

Method: We evaluate the accuracy of stretch on RFID Tattoos via a rotary system [4] which can control the level of stretch with a resolution of 5 μm. We manually clamp the two ends of the tattoo ensuring that it experiences exactly zero stretch force (just in the rest) when the stretcher shows zero on the scale. We then use the rotary system to stretch the RFID tattoo from 0 (in the resting state) to 30 mm in steps of 1mm. We repeat the experiments many times in different settings by placing the RFID reader in varying orientations with respect to the tattoo to emulate multipath and non-line-of-sight. As we stretch the stretchable RFID tattoos, we also co-locate our non-stretchable tattoos with the stretchable one, in order to isolate the stretch effect from the radio environment impact. We model the expected power by monitoring the response from each chip, where each RFID chip transmits its own unique pre-defined identification number following the standard RFID protocol. Further, we map the power to stretch by leveraging the ability of the tag to resonate at multiple discrete frequencies (see Sec. 4).

Results: We stretch the RFID tattoo from 0 to 30 mm. Fig. 10(a) shows the accuracy of stretch inference. We notice that RFTattoo achieves a 1.2 mm accuracy in stretch at the RSSI range of -40 to -45 dBm at around 30 cm distance between tattoo and reader. Further, our system drops in accuracy to 1.9 mm accuracy at lower RSSI (below -50 dBm) at around 1.2 m. We evaluate the tattoo's stretchability no further than 30 mm, since common facial movement produces an average stretch of 25 mm. Our RFID reader can successfully decode the response from the RFTattoo tag at a minimum of -72 dBm of RSSI which supports RFTattoo tags to be read from up to 2.5 meters away. During the experiment, we notice that as the RSSI reduces owing to the range or poor orientation, the inference accuracy drops. Since our inference model depends highly on power monitoring, we notice that a further-away tag may not respond at some resonant frequencies. One could achieve higher range performance by optimizing the material used for the RFID tags or using multiple RFID readers.

8.2 Tongue Positioning Accuracy

Method: RFTattoo uses tongue positions to distinguish between certain phonemes that are produced with the same facial gesture. We classify tongue positions using the resonant frequency shift of the RFID tattoo due to the near-field effects induced by the changes in tattoo's proximate environment, i.e., the tongue, as described in Sec. 4.2. We test this by classifying five tongue positions: (1) normal resting position, (2) upper jaw front: the tip of the tongue touching the base of the teeth on the upper jaw, (3) upper jaw back: the tip of the tongue touching the upper jaw ceiling, (4) lower jaw front: the tip of the tongue touching the base of the teeth on the

Table 2. Feature Set used by RFTattoo.

Minimal Features	Specialized Features
Mean, Variance, SD	Sample Entropy
Length	FFT Coefficients
Max, Min	Aggregated FFT
Longest strike above/below mean	Wavelet Transform coefficients
Variance > SD	Energy ratio by chunks
Mean absolute change	Linear trend (absolute value, slope)
Skewness	Time series complexity
% of reoccurring values	Quantile (5%,15%,85%,95%)

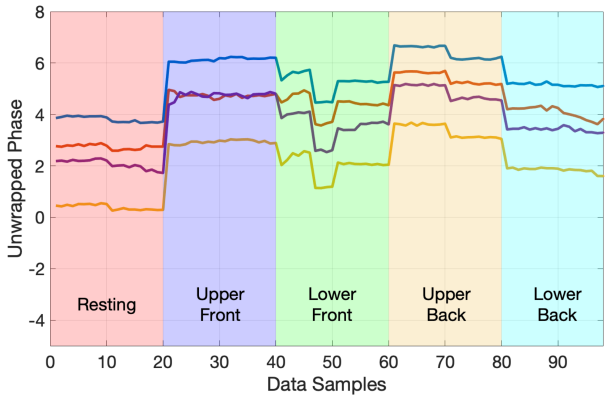

Fig. 9. Unwrapped phase of the reflected signal from four RFID chips in different RFID tattoos. The phase relationship changes with different tongue positions due to the impedance variation. We feed this as one of the raw data into feature extraction. See details in Sec. 4.2

lower jaw, and (5) lower jaw back: the tongue rolled inwards with the tip touching the lower jaw floor. Fig. 10(b) visualizes the different tongue positions. Our experiments were performed by requesting the users to keep the mouth slightly open allowing enough room for the tongue movement.

Results: Fig. 10(b) shows the confusion matrix for the Random Forest Classifier achieving an accuracy of 92%. Adaboost and Naive Bayes classifier also achieved similar accuracy on average. We have seen that our system is less sensitive to distinguish the position between the upper and lower jaw front, only one tag on both upper and lower jaw may be the reason to have lower accuracy. Fig. 9 shows the phase of the signal from different RFID chips while the tongue is moving among five tongue positions. We observe that the tongue can significantly influence the signal due to the impedance variation introduced to the RFID tattoos.

8.3 Accuracy of Viseme Classification

Method: To classify visemes, we collected an extensive set of RFID tattoo responses corresponding to the 38 phoneme classes from our test subjects. Remember that multiple phoneme classes belong to a single viseme class. In particular, the 38 phonemes can be grouped into 11 viseme classes. Our initial classification model was built based on the phoneme classes, but we realized that the phonemes belonging to the same viseme class were difficult to differentiate for reasons discussed in Sec. 6.1. We then decided to use the best representative phoneme from for each viseme class and build our classifier based on them. This is because once we have a viseme class, we can rely on our natural language processing framework to produce the word and sentence by feeding the likelihood of the candidate visemes to it. The details of the classifier are described in the Sec. 6. We ensured that our data set contained a mixture of responses corresponding to both facial gestures accompanied by sound and just the facial gestures by having the test subjects to utter the viseme without generating any sound.

Result: We trained our system on 9 classification models (see Sec.6.1) and select the best five classifier (shown in Fig.11(a)). Our Random Forests model gives the best performance with an average accuracy of 90% followed by Naive Bayes (84%) and Decision Tree (72%). Initial results based on phoneme classification revealed that it is difficult to differentiate between phonemes belonging to the same viseme solely based in the time and frequency response of the RFID tattoos. This is due to the inherent similarity of the phonemes belonging to the particular viseme class, that is, these phoneme sounds are generated with the same facial gesture and tongue position,

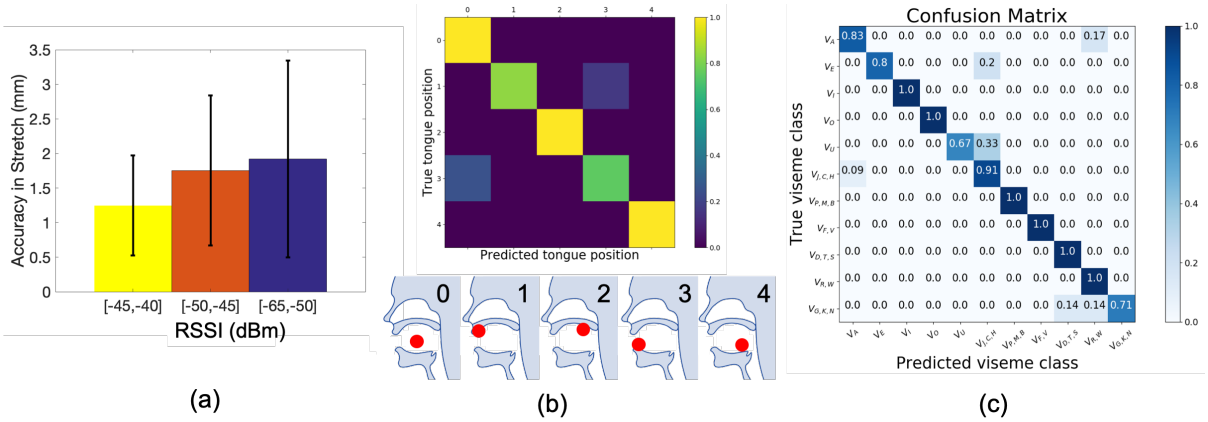

Fig. 10. (a) Stretch Accuracy: RFTattoo achieves a 1.4 mm median accuracy in stretch inference (Facial muscle movement has an average stretch of 25 mm.) (b) Tongue positioning: RFTattoo achieves 92% mean accuracy in classification of five tongue positioning. Index 0 to 4 represent tongue positions at: normal resting position, upper jaw front, upper jaw back, lower jaw front, lower jaw back. (c) Confusion matrix of 11 visemes: RFTattoo achieves an average test accuracy of 90%.

there by producing almost the similar amount of stretch and impedance change in the RFID tattoos. This led to a low classification accuracy when classifying phonemes mapping to the same viseme for each of the models, albeit a good accuracy classifying phonemes mapping to different visemes. We then chose the best representative phoneme for each of the viseme class and trained the classifier based on them. Fig. 10(c) shows the confusion matrix for the Random Forest Classifier based on the dataset collected across users. While almost all the viseme classes can be classified with sufficiently large accuracy, class V_U (corresponding to the "u" sound) has a low accuracy due to the nature of the stretch it produces on different face structures. While at first glance, it would appear that the Viseme class V_U will produce the most distinct response as it generates the maximum facial distortion, upon taking a closer look we find that this distortion is highly dependent on the user's specific face structure. This is not an issue with other viseme classes as the facial distortion is generally consistent across users with diverse face structures. Another observation from our results is that the classifier performance is virtually unchanged by the presence or absence of sound.

8.4 Accuracy in the Word Classification

Method: Although the viseme classification produced a substantially high accuracy, we observed that mapping a combinations of visemes to a word purely by stitching together the responses of different viseme candidates proved to be relatively unreliable. This is partly because the probability of an error in the word increases exponentially with the number of phonemes it is composed of. However, such an approach would ignore the structure of the English language, where certain words are more common than others and certain combinations of phonemes are valid while others are not. This led us to building classifier system at the word-level to feed the candidate word likelihood to the NLP layer. Our vocabulary dataset is composed of 100 most commonly used words in English. To evaluate the impact of the size of vocabulary on the accuracy, we randomly select 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 words in our vocabulary dataset and follow our dataset management to split the words into non-overlapping training and testing datasets. We use the same model feature configuration to train all the models for different sizes of vocabulary. We notice that different words may have varying accuracy that impacts our system, so we repeat the experiment over 100 iterations for each size of the vocabulary. Apart from the most commonly used words, we also incorporate a good mix of monosyllabic, disyllabic and tri-syllabic words to

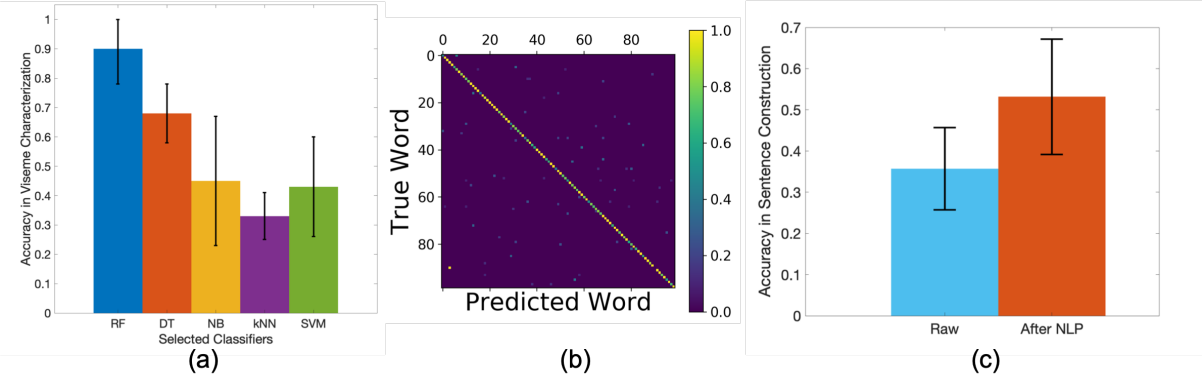

Fig. 11. (a) Accuracy in predicting viseme classes for selected classifiers: Random Forest, Decision Tree, Naive Bayes, kNN and SVM. (b) Confusion matrix of 100 words classification. RFTattoo achieves an average test accuracy of 86%. (c) Accuracy in sentence construction.

train the classification model. As in the tests for viseme classification, we ensured that the dataset contained the responses corresponding to facial expressions accompanied with sound and without sound.

Result: We trained our classifier based on the 10 classifiers listed in the previous section. We achieve an average test accuracy of 90% based on Random Forest classifier followed by 67% on Decision trees with a size of 50 words vocabulary dataset. We achieve an average accuracy of 86% with a size of 100 words. Fig. 11(b) shows the confusion matrix obtained on one of the test dataset containing a good mix of all the 100 words uttered in different ways (with/without generating sound). We observe that one word is nearly recognized to be a different word, that is because of the lack of speech context, which led to our NLP solution. It should also be noted that a good proportion of words can be accurately classified since the utterance of a word generates a much richer time and frequency response owing to the longer time series containing instances of multiple stretches. Fig. 12(a) shows our system performance versus the size of vocabulary from 10 to 100 words with a step size of 10. We observe that our system accuracy drops when the vocabulary size gets larger. One possible solution to that is to use more RFID chips that resonate at more fine-grained discrete frequencies to model the stretch.

The Effect of Misalignment: We have tested our system on previously untrained speakers, where no pre-calibration is performed before the experiments. We observe a 72% accuracy over the above 50 words using our trained model based on Random Forest classifier.

Different Types of Users: We achieve an average accuracy of 86% among 10 users. Participants who have temporary dysphonia have a very similar accuracy of 85%. During the experiments, we also observe an average accuracy of 87.1% and 86.9% for participants speaking with sound and without sound.

Number of Reader Antennas: We conduct an experiment to evaluate the impact of the number of RFID reader antennas. Specifically, we collect 20 words using a single antenna and four antennas connected to the RFID reader. The four antennas are arranged in line with 5 cm spacing. Again, we follow our dataset management mechanism to obtain the training and testing dataset (described in Sec. 7). In Fig. 12(c), we show the prediction accuracy versus the number of reader antennas, we got 93.1% and 97.2% accuracy when using a single antenna and four antennas. To process the data from four antennas, we individually feed raw data from different antennas into the prediction model and average the accuracy across all the antennas. We observe a higher accuracy when using four antennas since multiple antennas give us richer spatial diversity in the RFID tags reading (higher overall

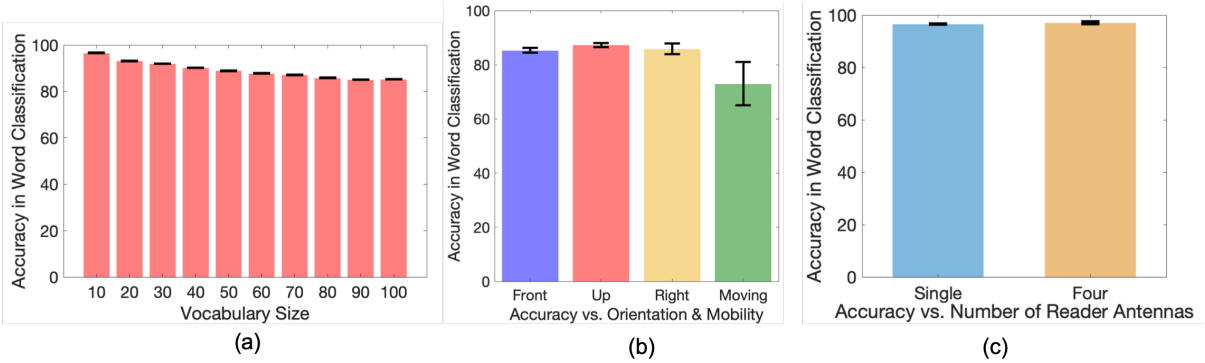

Fig. 12. Accuracy (%) vs. (a) Vocabulary Size; (b) user's orientation; (c) number of reader antennas.

response rate of the tags and better spatial resolution). In the rest of our evaluation, we obtain results using a single reader antenna, given that it allows for more portable and cost-effective platforms.

The Effect of User Orientation: We ask users to face the reader antenna with three different orientations: front, top and right. We collect 20 words for each orientation and use them as testing data. In Fig. 12(b), we show that our system is robust at most of the orientations. While we get 85.3%, 87.2% and 85.8% towards different orientations, we also observe that at some orientations, the accuracy dips when multiple RFID tattoos experience extreme shadowing from the human body.

Mobility / Body Position: We conduct an experiment where users are moving with walking speed (*sim* 1 m/s) and the relative position between the tattoos and the reader antenna changes over time. We collect 20 words and use them as testing data and predict the word accuracy using our trained model of 100 words. We observe a 76.5% accuracy while the users are moving at walking speed versus accuracy of 86% when the user is static.

8.5 Accuracy in Sentence Construction

Method: We conducted a pilot experiment for the sentence construction in a regular office space. We first request participants to wear four RFID tags in a way they feel comfortable. We then asked participants to speak three warm-up words to get familiar with the system. We also perform a calibration phase per-user (see Sec. 7). During the experiment, we presented each participant a paper sheet and asked them to read the sentences on the sheet without sound, 10 repetitions for each sentences (to measure system variance). We performed two types of experiments: (1) First, the sentences on the sheet were among twenty candidate sentences often used in daily conversations with the pool of sentences known to RFTattoo a priori. (2) Second, the sentences are arbitrary chosen grammatically correct sentences that span 4 words to 10 words, the typical range of sentence lengths in English speech [58], with the pool of sentences spoken by the user not known a priori to RFTattoo. At the end of the experiment, we helped the participants remove the RFID tags and provided alcohol-soaked cotton swab to help participants clean the glue residue.

Result: Figure 11(c) illustrates the sentence level recognition accuracy of RFTattoo. We first observe that for the 20 commonly used sentences in day-to-day use known to our system, RFTattoo shows an average of 91% accuracy. In contrast, the raw recognition accuracy for sentences unseen by RFTattoo is 35.7%. Integrating natural language processing based correction further boosts the average accuracy to 53.2%. We note that this is within the performance range for unknown sentences of state-of-the-art vision-based lip-reading software that requires line-of-sight (e.g. 46.8% in [22]). We also find that RFTattoo works better for longer sentences, which contains more contextual information. Our results reveal that RFTattoo holds promise in reconstructing sentences for

users with voice impairments. Our accuracy can be further improved over time with more data to tune to the user's particular speaking habits.

9 LIMITATIONS

Calibration: RFTattoo requires a light-weight calibration step to be performed a priori. To achieve optimal system performance, one must re-calibrate should RFTattoo tags be peeled off and on, or with natural wear. We describe the details of the procedure in Sec. 7

Extreme shadowing: Due to the 2-D dipole antenna used by RFID tags, RFTattoo tags may experience shadowing due to the human body at certain orientations of the tag relative to the reader. This can in part be circumvented by attaching more RFID tattoos to the user's face.

Words out of the vocabulary dataset: RFTattoo has poor accuracy in predicting the untrained words. This is a common problem shared by voice recognition systems as well as lip-reading systems.

Sensing and Computational Latency: Our current RFTattoo prototype has an end-to-end refresh rate of around 0.8 Hz which is 48 words per minute, running on MacBook Pro. The window size for RFTattoo feature extraction is set to be 1 second. On average, the stretch inference and feature extraction takes 0.24 second in total, the output of our prediction model for a 4-word sentences takes 5.2 milliseconds. The Impinj RFID reader can process 3000 readings per second. We note that our end-to-end latency can be improved with better compute infrastructure at the edge.

Communication Range: RFTattoo has a limited communication range up to 2.5 meters. The reason is two-fold. First, human skin has lossy dielectric material properties, which means the range performance of the RFTattoo tags are highly associated with the permittivity and conductivity of the skin. Further, we note that the electrical properties of the human body can vary from person to person. RFTattoo shows excellent performance when the RFID reader is around the waist. Second, RFTattoo infers the tag stretch based on the response of the different resonant chips, and its range is a function of the choice of material of the tag. One could achieve a higher range performance by optimizing the material used for the RFID tags or using blind beamforming techniques [73].

10 DISCUSSION

Facial Expressions: Many other facial movements other than talking can also affect the face, such as expressions (e.g., a smile), eating and drinking. This is also a well-known challenge for visual lip-reading techniques. Indeed, recognizing the type of facial expression itself remains a challenge for lip-reading systems [61]. Past work has relied on specialized features such as the movement of lip corners to distinguish between similar expressions, e.g., posed versus genuine smiles [31]. While the scope of RFTattoo is restricted to speech recognition, designing RFID tattoos which can perform advanced expression sensing by sensing the stretch of lip corners and beyond remains an open challenge for the future work.

Interference in the 915 MHz Band: We test our system where only one RFID reader is present in the environment. Our system is at least as resilient as standard RFID systems since we use commercial RFID chips within our RFID tattoos. As with any wireless system, the presence of interfering sources in the shared 915 MHz band can cause a reduction in the performance of RFTattoo. We believe future implementations of RFTattoo can benefit from the rich literature on RFID interference cancellation [69].

RFID Tattoo Signal Collisions: RFTattoo uses commercial RFID chips which fully run the standard commercial RFID protocol (EPC Gen2 UHF RFID protocol). This protocol inherently mitigates collisions due to the presence of a large number of RFID tags based on Slotted Aloha. Indeed, the signal collision gets worse if the environment

presents massive RFID tags, like the warehouses. One potential method is to pose SELECT queries [15] on RFID readers to which only a user-defined group of RFID tags respond to the RFID reader. For now, our RFID reader can process 100 RFID tags simultaneously with 3000 tag readings per second.

RFTattoo System Cost: The cost of each RFID tattoo is similar to the commercial RFID tag. While the RFID reader used in our experiment cost around $1500, there are cheaper RFID readers ($200) available in the market. We envision that RFID reader chips will be integrated with personal devices such as smartphones to preserve user privacy.

11 CONCLUSION

This paper presents, to our knowledge, the first system that recognizes intended speech of users with voice impairments using light-weight RFID tattoos attached to the face. We present algorithms that recognize subtle stretches of the tattoos as well as movement of the tongue by processing signals reflected off the tags received at a handheld RFID reader. Our system then builds a natural language processing framework that recognizes various facial gestures associated with speech to construct meaningful words and sentences. We present results from a detailed user study that reveals the promise of our approach in recognizing intended speech, even when users do not make any sounds.

ACKNOWLEDGMENTS

We thank the members of the WiTech group for their insightful discussions and anonymous reviewers for their constructive feedback. We would like to thank NSF (grants 1718435, 1657318, 1823235 and 1837607) for their support and all the volunteers for participating in our study.

REFERENCES

[1] 2018. Impinj RFID Reader - ItemSense Reference Guide. https://developer.impinj.com/itemsense/docs/reference-guide. (12 2018).
[2] 2018. The OEC: Facts about the language. https://web.archive.org/web/20111226085859/http://oxforddictionaries.com/words/the-oec-facts-about-the-language. (12 2018).
[3] 2018. TextSpeak. https://www.alimed.com/textspeak.html. (12 2018).
[4] 2018. Velmex Positioning Products. https://www.velmex.com/Products/index.html. (12 2018).
[5] 2018. Voice Disorder - an overview | SicenceDirect Topics. https://www.sciencedirect.com/topics/medicine-and-dentistry/voice-disorder. (12 2018).
[6] 2019. How a new technology is changing the lives of people who cannot speak. https://www.theguardian.com/news/2018/jan/23/voice-replacement-technology-adaptive-alternative-communication-vocalid. (07 2019).
[7] 2019. How Siri Works. https://electronics.howstuffworks.com/gadgets/high-tech-gadgets/siri4.html. (01 2019).
[8] 2019. Murata Electronics. https://www.mouser.com/murataelectronics/. (01 2019).
[9] 2019. Nu-Vois III Artificial Larynx. http://www.weitbrecht.com/NuVois-III-speech-aid.html. (01 2019).
[10] 2019. Wolf paulus' Journal: Viseme Model with 12 mouth shapes. https://wolfpaulus.com/jounal/software/lipsynchronization/. (01 2019).
[11] 2019. Xerafy Dot-on. https://docs.wixstatic.com/ugd/9b3bd3_6abffdaa1ce94028b4ff2f0c40020ce7.pdf. (01 2019).
[12] Shreya Agrawal, Verma Rahul Omprakash, et al. 2016. Lip reading techniques: A survey. In *Applied and Theoretical Computing and Communication Technology (iCATccT), 2016 2nd International Conference on*. IEEE, 753–757.
[13] HFSS Ansoft. 2011. Ver. 13. *Ansoft Corporation* (2011).
[14] Yannis M. Assael, Brendan Shillingford, Shimon Whiteson, and Nando de Freitas. 2016. LipNet: Sentence-level Lipreading. *CoRR* abs/1611.01599 (2016). arXiv:1611.01599 http://arxiv.org/abs/1611.01599
[15] Henri Barthel. 2005. EPCglobal–RFID standards & regulations. (2005).
[16] Helen L. Bear. 2017. Decoding visemes: improving machine lipreading (PhD thesis). *CoRR* abs/1710.01288 (2017). arXiv:1710.01288 http://arxiv.org/abs/1710.01288
[17] Abdelkareem Bedri, Himanshu Sahni, Pavleen Thukral, Thad Starner, David Byrd, Peter Presti, Gabriel Reyes, Maysam Ghovanloo, and Zehua Guo. 2015. Toward silent-speech control of consumer wearables. *Computer* 48, 10 (2015), 54–62.
[18] Bradley J Betts and Charles Jorgensen. 2005. Small vocabulary recognition using surface electromyography in an acoustically harsh environment. (2005).

[19] P. Bhartia, K. V. S. Rao, and R. S. Tomar. 1991. *Millimeter-wave microstrip and printed circuit antennas / P. Bhartia, K.V.S. Rao, R.S. Tomar.* Artech House Boston. xiii, 322 p. : pages.

[20] Katharine Brigham and BVK Vijaya Kumar. 2010. Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy. In *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on*. IEEE, 1–4.

[21] Luca Catarinucci, Riccardo Colella, and Luciano Tarricone. 2012. Smart prototyping techniques for UHF RFID tags: electromagnetic characterization and comparison with traditional approaches. *Progress in Electromagnetics Research* 132 (2012), 91–111.

[22] Joon Son Chung, Andrew W Senior, Oriol Vinyals, and Andrew Zisserman. 2017. Lip Reading Sentences in the Wild.. In *CVPR*. 3444–3453.

[23] Martin Cooke, Jon Barker, Stuart Cunningham, and Xu Shao. 2006. An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America* 120, 5 (2006), 2421–2424.

[24] Filippo Costa, Simone Genovesi, and Agostino Monorchio. 2013. Chipless RFID trasponders by using multi-resonant High-Impedance Surfaces. *2013 International Symposium on Electromagnetic Theory* (2013), 401–403.

[25] Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*. Association for Computational Linguistics, 76–87.

[26] Nicolas EVENO, Patrice DELMAS, and Pierre-Yves COULON. [n. d.]. VERS LâĂŹEXTRACTION AUTOMATIQUE DES LÈVRES DâĂŹUN VISAGE PARLANT. ([n. d.]).

[27] Cletus G Fisher. 1968. Confusions among visually perceived consonants. *Journal of Speech and Hearing Research* 11, 4 (1968), 796–804.

[28] Masaaki Fukumoto. 2018. SilentVoice: Unnoticeable Voice Input by Ingressive Speech. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. ACM, New York, NY, USA, 237–246. https://doi.org/10.1145/3242587.3242603

[29] Jinlan Gao, Johan Siden, and Hans-Erik Nilsson. 2011. Printed electromagnetic coupler with an embedded moisture sensor for ordinary passive RFID tags. *IEEE Electron Device Letters* 32, 12 (2011), 1767–1769.

[30] CMU Speech Group. 2019. The CMU Pronouncing Dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict. (2019). (Accessed on 01/17/2019).

[31] Hui Guo, Xiao-Hui Zhang, Jun Liang, and Wen-Jing Yan. 2018. The dynamic features of lip corners in genuine and posed smiles. *Frontiers in psychology* 9 (2018), 202.

[32] Unsoo Ha, Yunfei Ma, Zexuan Zhong, Tzu-Ming Hsu, and Fadel Adib. 2018. Learning Food Quality and Safety from Wireless Stickers. In *Proceedings of the 17th ACM Workshop on Hot Topics in Networks (HotNets '18)*. ACM, New York, NY, USA, 106–112. https://doi.org/10.1145/3286062.3286078

[33] Ahmad Basheer Hassanat. 2014. Visual words for automatic lip-reading. *arXiv preprint arXiv:1409.6689* (2014).

[34] Tess Hellebrekers, Kadri Bugra Ozutemiz, Jessica Yin, and Carmel Majidi. 2018. Liquid Metal-Microelectronics Integration for a Sensorized Soft Robot Skin. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5924–5929.

[35] Xiaonan Huang, Kitty Kumar, Mohammad K Jawed, Amir M Nasab, Zisheng Ye, Wanliang Shan, and Carmel Majidi. 2018. Chasing biomimetic locomotion speeds: Creating untethered soft robots with shape memory alloy actuators. *Science Robotics* 3, 25 (2018), eaau7557.

[36] DAVIDR Jackson and NICOLAOSG Alexopoulos. 1986. Analysis of planar strip geometries in a substrate-superstrate configuration. *IEEE transactions on antennas and propagation* 34, 12 (1986), 1430–1438.

[37] Kyung-In Jang, Sang Youn Han, Sheng Xu, Kyle E Mathewson, Yihui Zhang, Jae-Woong Jeong, Gwang-Tae Kim, R Chad Webb, Jung Woo Lee, Thomas J Dawidczyk, et al. 2014. Rugged and breathable forms of stretchable electronics with adherent composite substrates for transcutaneous monitoring. *Nature communications* 5 (2014), 4779.

[38] Matthias Janke and Lorenz Diener. 2017. EMG-to-speech: Direct generation of speech from facial electromyographic signals. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)* 25, 12 (2017), 2375–2385.

[39] Haojian Jin, Jingxian Wang, Zhijian Yang, Swarun Kumar, and Jason Hong. 2018. Rf-wear: Towards wearable everyday skeleton tracking using passive rfids. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. ACM, 369–372.

[40] Haojian Jin, Jingxian Wang, Zhijian Yang, Swarun Kumar, and Jason Hong. 2018. WiSh: Towards a Wireless Shape-aware World Using Passive RFIDs. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '18)*. ACM, New York, NY, USA, 428–441. https://doi.org/10.1145/3210240.3210328

[41] Hsin-Liu (Cindy) Kao, Christian Holz, Asta Roseway, Andres Calvo, and Chris Schmandt. 2016. DuoSkin: Rapidly Prototyping On-skin User Interfaces Using Skin-friendly Materials. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. ACM, New York, NY, USA, 16–23. https://doi.org/10.1145/2971763.2971777

[42] Arnav Kapur, Shreyas Kapur, and Pattie Maes. 2018. AlterEgo: A Personalized Wearable Silent Speech Interface. In *23rd International Conference on Intelligent User Interfaces*. ACM, 43–53.

[43] Jiseok Kim, Zheng Wang, and Woo Soo Kim. 2014. Stretchable RFID for wireless strain sensing with silver nano ink. *IEEE Sensors Journal* 14, 12 (2014), 4395–4401.

[44] Alexandros Koumparoulis, Gerasimos Potamianos, Youssef Mroueh, and Steven J Rennie. [n. d.]. Exploring ROI size in deep learning based lipreading. ([n. d.]).
[45] Alex Krizhevsky et al. 2009. *Learning multiple layers of features from tiny images*. Technical Report. Citeseer.
[46] Taoran Le, Ryan A Bahr, Manos M Tentzeris, Bo Song, and Ching-ping Wong. 2015. A novel chipless RFID-based stretchable and wearable hand gesture sensor. In *Microwave Conference (EuMC), 2015 European*. IEEE, 371–374.
[47] Yann LeCun. [n. d.]. The MNIST database of handwritten digits. *http://yann. lecun. com/exdb/mnist/* ([n. d.]).
[48] Soonkyu Lee and DongSuk Yook. 2002. Audio-to-visual conversion using hidden markov models. In *Pacific Rim International Conference on Artificial Intelligence*. Springer, 563–570.
[49] Ruibo Liu, Qijia Shao, Siqi Wang, Christina Ru, Devin Balkcom, and Xia Zhou. 2019. Reconstructing Human Joint Motion with Computational Fabrics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 19 (March 2019), 26 pages. https://doi.org/10.1145/3314406
[50] Yuhao Liu, Matt Pharr, and Giovanni Antonio Salvatore. 2017. Lab-on-skin: a review of flexible and stretchable electronics for wearable health monitoring. *ACS nano* 11, 10 (2017), 9614–9635.
[51] Joanne Lo, Doris Jung Lin Lee, Nathan Wong, David Bui, and Eric Paulos. 2016. Skintillates: Designing and creating epidermal interactions. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM, 853–864.
[52] Yunfei Ma, Nicholas Selby, and Fadel Adib. 2017. Minding the billions: Ultra-wideband localization for deployed rfid tags. In *ACM MobiCom*.
[53] Hiroyuki Manabe, Akira Hiraiwa, and Toshiaki Sugimura. 2003. "Unvoiced Speech Recognition Using EMG - Mime Speech Recognition". In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 794–795. https://doi.org/10.1145/765891.765996
[54] Eric Markvicka, Guanyun Wang, Yi-Chin Lee, Gierad Laput, Carmel Majidi, and Lining Yao. 2019. ElectroDermis: Fully Untethered, Stretchable, and Highly-Customizable Electronic Bandages. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 632, 10 pages. https://doi.org/10.1145/3290605.3300862
[55] Eric J Markvicka, Michael D Bartlett, Xiaonan Huang, and Carmel Majidi. 2018. An autonomously electrically self-healing liquid metal–elastomer composite for robust soft-matter robotics and electronics. *Nature materials* (2018), 1.
[56] Lindasalwa Muda, Mumtaj Begam, and Irraivan Elamvazuthi. 2010. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *arXiv preprint arXiv:1003.4083* (2010).
[57] NASA. 2004. Subvocal Speech. https://www.nasa.gov/centers/ames/news/releases/2004/subvocal/subvocal.html. (2004).
[58] Margaret Morse Nice. 1925. Length of sentences as a criterion of a childâĂŹs progress in speech. *Journal of Educational Psychology* 16 (1925), 370–379.
[59] C Occhiuzzi, C Paggi, and G Marrocco. 2011. Passive RFID strain-sensor based on meander-line antennas. *IEEE Transactions on Antennas and Propagation* 59, 12 (2011), 4836–4840.
[60] Dumtoochukwu O Oyeka, John C Batchelor, and Ali Mohamad Ziai. 2017. Effect of skin dielectric properties on the read range of epidermal ultra-high frequency radio-frequency identification tags. *Healthcare technology letters* 4, 2 (2017), 78–81.
[61] Maja Pantic and Ioannis Patras. 2006. Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 36, 2 (2006), 433–449.
[62] Swadhin Pradhan, Eugene Chai, Karthikeyan Sundaresan, Lili Qiu, Mohammad A. Khojastepour, and Sampath Rangarajan. 2017. RIO: A Pervasive RFID-based Touch Gesture Interface. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (MobiCom '17)*. ACM, New York, NY, USA, 261–274. https://doi.org/10.1145/3117811.3117818
[63] Robert F Shepherd, Filip Ilievski, Wonjae Choi, Stephen A Morin, Adam A Stokes, Aaron D Mazzeo, Xin Chen, Michael Wang, and George M Whitesides. 2011. Multigait soft robot. *Proceedings of the national academy of sciences* 108, 51 (2011), 20400–20403.
[64] H-W Son and C-S Pyo. 2005. Design of RFID tag antennas using an inductively coupled feed. *Electronics Letters* 41, 18 (2005), 994–996.
[65] George Sterpu, Christian Saam, and Naomi Harte. 2018. Attention-based Audio-Visual Fusion for Robust Automatic Speech Recognition. (2018). https://doi.org/10.1145/3242969.3243014 arXiv:arXiv:1809.01728
[66] Patrick Suppes, Zhong-Lin Lu, and Bing Han. 1997. Brain wave recognition of words. *Proceedings of the National Academy of Sciences* 94, 26 (1997), 14965–14969.
[67] U Tata, H Huang, RL Carter, and JC Chiao. 2008. Exploiting a patch antenna for strain measurements. *Measurement Science and Technology* 20, 1 (2008), 015201.
[68] Lijun Teng, Kewen Pan, Markus P Nemitz, Rui Song, Zhirun Hu, and Adam A Stokes. 2018. Soft Radio-Frequency Identification Sensors: Wireless Long-Range Strain Sensors Using Radio-Frequency Identification. *Soft robotics* (2018).
[69] Michael Thomas, William Colleran, Erik Fountain, and Todd Humes. 2006. Interference cancellation in RFID systems. (May 11 2006). US Patent App. 10/981,893.
[70] Guanhua Wang, Yongpan Zou, Zimu Zhou, Kaishun Wu, and Lionel M Ni. 2016. We can hear you with wi-fi! *IEEE Transactions on Mobile Computing* 15, 11 (2016), 2907–2920.

[71] Ju Wang, Omid Abari, and Srinivasan Keshav. 2018. Challenge: RFID Hacking for Fun and Profit. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom '18)*. ACM, New York, NY, USA, 461–470. https://doi.org/10.1145/3241539.3241561

[72] Ju Wang, Jie Xiong, Xiaojiang Chen, Hongbo Jiang, Rajesh Krishna Balan, and Dingyi Fang. 2017. TagScan: Simultaneous Target Imaging and Material Identification with Commodity RFID Devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (MobiCom '17)*. ACM, New York, NY, USA, 288–300. https://doi.org/10.1145/3117811.3117830

[73] Jingxian Wang, Junbo Zhang, Rajarshi Saha, Haojian Jin, and Swarun Kumar. 2019. Pushing the Range Limits of Commercial Passive RFIDs. In *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19)*. USENIX Association, Boston, MA, 301–316. https://www.usenix.org/conference/nsdi19/presentation/wangjingxian

[74] Martin Weigel, Tong Lu, Gilles Bailly, Antti Oulasvirta, Carmel Majidi, and Jürgen Steimle. 2015. Iskin: flexible, stretchable and visually customizable on-body touch sensors for mobile computing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2991–3000.

[75] Martin Weigel, Aditya Shekhar Nittala, Alex Olwal, and Jürgen Steimle. 2017. Skinmarks: Enabling interactions on body landmarks using conformal skin electronics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3095–3105.

[76] Anusha Withana, Daniel Groeger, and Jürgen Steimle. 2018. Tacttoo: A thin and feel-through tattoo for on-skin tactile output. In *The 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, 365–378.

[77] Lei Yang, Yekui Chen, Xiang-Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. 2014. Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 237–248.

[78] Daniel J Yeager, Jeremy Holleman, Richa Prasad, Joshua R Smith, and Brian P Otis. 2009. Neuralwisp: A wirelessly powered neural interface with 1-m range. *IEEE Transactions on Biomedical Circuits and Systems* 3, 6 (2009), 379–387.

[79] Jun Yin, Jun Yi, Man Kay Law, Yunxiao Ling, Man Chiu Lee, Kwok Ping Ng, Bo Gao, Howard C Luong, Amine Bermak, Mansun Chan, et al. 2010. A system-on-chip EPC Gen-2 passive UHF RFID tag with embedded temperature sensor. *IEEE Journal of Solid-State Circuits* 45, 11 (2010), 2404–2420.

[80] Victoria Young and Alex Mihailidis. 2010. Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review. *Assistive Technology* 22, 2 (2010), 99–112.